1970

Survey designs utilizing prior information

Cary Tsuguo Isaki
Iowa State University

71-14,232

ISAKI, Cary Tsuguo, 1944-
SURVEY DESIGNS UTILIZING PRIOR
INFORMATION.

Iowa State University, Ph.D., 1970
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

SURVEY DESIGNS UTILIZING PRIOR INFORMATION

by

Cary Tsuguo Isaki

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Ames, Iowa

1970

TABLE OF CONTENTS

I. INTRODUCTION

To consider the general problem of inference for a finite population we first introduce suitable notation. The elements of the finite population are denoted by $U = \{U_i: \; i=1,2,\ldots N\}$, the parameter space is denoted by Θ, the set of possible actions is denoted by A, the set of possible probability systems is denoted by $\underline{P}$ and the prior information is denoted by I.

The parameter space Θ is the N-dimensional vector space, $R^N, \Theta = \{\theta = (y_1, y_2, \ldots, y_N) : y_i \varepsilon R^1, i=1,2,\ldots N\}$ where y_i is the characteristic associated with unit U_i. Inference is usually for some function of the parameter. Such functions include the total, the mean and the mean of those y_i's greater than some specified number C.

The set of possible actions A, is the set of estimators $a \varepsilon A$, of the parameter of interest. Denote by S, the collection of all groupings of elements U_i (by a grouping we shall mean any one of the possible configurations obtainable from U). For example a grouping may be $s_3 = (U_1, U_3, U_{18})$. We exclude the null set and the whole population U. The estimator a is a function of the elements of the parameter $(y_1, y_2, \ldots y_N)$ associated with the elements contained in an element $s \varepsilon S$ and of the prior information I.

The set of probability systems $\underset{\sim}{P}$ represents all possible methods of selecting samples (groupings) s from S. Nonnegative weights, p(s), are assigned to s such that

$$\sum_{s \varepsilon S} p(s) = 1 \quad .$$

Denoting an index set by $\mathcal{A}$,

$$\underset{\sim}{P} = \left\{ p_\alpha : \sum_{s \varepsilon S} p_\alpha(s) = 1\right.$$

and

$$\left. p_\alpha(s) \geq 0 \ \forall \ s\varepsilon S,\ \alpha \varepsilon \mathcal{A} \right\} \quad .$$

The symbol p_α will be called a sampling design and we will henceforth drop the subscript α. We now define what we shall mean by a survey design.

Def. A survey design consists of a sampling design and an estimator for the population characteristic.

The prior information I includes all possible information available to the survey designer. Naturally, it is assumed that the designer does not know the actual parameter $(y_1, y_2, \ldots y_N)$. For example, prior information may be in the form of an auxiliary variable $(x_1, \ldots x_N)$ which is correlated with $(y_1, \ldots y_N)$. Or it may be possible to stratify (group) the units of U of the population sc that the y_i in any particular group possess nearly the same value.

The prior information will influence the survey designer's choice of action and of sampling design. He utilizes the available information, I, in selecting the best, in some sense, $a \varepsilon A$ and $p \varepsilon \underline{P}$. Let us denote by $W(a(s),f(\Theta))$ the loss incurred by the survey designer when he chooses $a \varepsilon A$, $f(\Theta)$ is the parameter of interest, and s is the sample selected. The loss function is assumed to be such that if $a(s) = f(\Theta)$, he incurs no loss. The loss function is also assumed to be bounded. For example, a loss function may be of the form $W(a(s),f(\Theta)) = (a(s)-f(\Theta))^2$ where $a(s)$ and $f(\Theta)$ are finite. Let

$$E_p[W(a(s),f(\Theta))] = \sum_{s \varepsilon S} W(a(s),f(\Theta))p(s)$$

denote the expected loss which is a function of a, p and $f(\Theta)$. Then $a \varepsilon A$ and $p \varepsilon \underline{P}$ may be said to be best with respect to the loss function $W(a(s),f(\Theta))$ if they jointly minimize the expected loss. When $W(a(s),f(\Theta)) = (a(s)-f(\Theta))^2$, the expected loss is the Mean Squared Error (M.S.E.).

The general problem of finding the "best" $a \varepsilon A$ and $p \varepsilon \underline{P}$ has been considered in the literature. Often the space A has been restricted to a subclass of the class of all possible estimators. Horwitz and Thompson (28), Godambe (16), Koop (31) and Roy and Chakravarti (50) restricted A to be a subclass of linear estimators. Hanurav (22) restricted A to the class of polynomial estimators. Often, construction of a suitable sampling design is of sole interest and emphasis is placed on obtaining a p which offers some desirable properties. For example, given a, Hanurav (24), Sampford (51), Midzuno (37) and Fuller (15) have developed unequal probability sampling designs for which nonnegative unbiased estimators of variance are available. Neyman (41) and Ericson (10) considered the optimal allocation of a sample to strata given that the population had previously been stratified and given that the stratified mean was to be used as the estimator. Dalenius (7) and Dalenius and Hodges (8) considered an "optimum" method of stratifying the population.

One group of authors have utilized the expected variance as a criterion of bestness. Cochran (4) first applied the expected variance in a survey design situation. Some other contributions in this area include the rejective sampling design of Hájek (20) and the minimax estimation considered by Aggarwal (1). Other authors who have used the expected variance are Godambe (16), Rao, Hartley and Cochran (48) and Hanurav (25). The expected variance will be treated in detail in the next section.

It appears that some restrictions on the general problem are required to obtain a satisfactory solution. Often the class of estimators is restricted to unbiased estimators. This is a common restriction imposed by many authors since unbiased estimators have the advantage of i) simple interpretation and ii) additivity. Other common restrictions include fixed sample size designs and admissible estimators.

In the following section several treatments of the problem of selecting the "best" $a \varepsilon A$ and $p \varepsilon \underline{P}$ will be discussed. Emphasis will be placed on the criterion used in deriving estimators or sampling designs.

II. DISCUSSION OF CRITERIA IN USE

A. Mean Square Error (M.S.E.)

It is natural to call the sample design and estimator which yields the smallest M.S.E. "best". The M.S.E. may be used to evaluate both the estimator and the sampling design given the prior information. Godambe (16) has shown that given the finite population $(y_1,\ldots,y_N)$ and if A is restricted to the class of linear unbiased estimators, and $\underline{P}$ is restricted to the class of sampling designs admitting unbiased estimators of the variance of $a \varepsilon A$, that there does not exist a uniformly (for all y) minimum variance unbiased estimator for $\bar{Y}_N$. More recently, Godambe and Joshi (17) have extended the result to the class of unbiased estimators.

Hanurav (25) has shown that if $\underline{P}$ is restricted to the class of Unicluster Designs, i.e. if for any two samples s_1 and s_2 such that $p(s_1) > 0$, $p(s_2) > 0$

i) s_1 and s_2 are either disjoint (no common elements)
or

ii) every unit $u_i \varepsilon s_1$ is also in s_2 and vice-versa,

then the Horwitz-Thompson estimator is the minimum variance unbiased estimator in the class of linear estimators (it is the only one). Of course, it is assumed that these Unicluster Designs admit an unbiased estimator. One drawback of the Unicluster Designs is that an unbiased estimate of variance of the estimator is not available. An example of a Unicluster Design is systematic sampling with equal probabilities.

B. Admissibility

Although an estimator is not minimum M.S.E. it may satisfy the weaker condition of admissibility. For a given sampling design p, an estimator e_1 is admissible for $f(\theta)$ if there does not exist another estimator e_2 such that M.S.E. $(e_2) \leq$ M.S.E. (e_1) uniformly for all $(y_1, \ldots, y_N)$ and M.S.E. $(e_2) <$ M.S.E. (e_1) for at least one particular $(y_1, \ldots, y_N)$. Admissibility restricts consideration to "reasonable estimators". Clearly, if we restrict ourselves to unbiased estimators, we may replace M.S.E. by the variance V. Since in the definition of admissibility, a sampling design p, is assumed, admissibility as defined does not provide a guide for the determination of the "best" sampling design. An estimator e which is admissible with respect to a sampling design p_1, may not be admissible with respect to another sampling design p_2.

Godambe and Joshi (17) have shown that if the design is such that the inclusion probability for every unit in the population is positive and if we restrict consideration to unbiased estimators, then the Horwitz-Thompson (H.T.) estimator is admissible. On the basis of the admissibility criterion, it would be reasonable to select the estimator and sampling design such that the estimator is admissible and to choose the sampling design which minimizes the M.S.E. of the estimator. If we restrict A to the class of unbiased estimators, we could consider the H.T. estimator as a likely candidate. The variance expression for the H.T. estimator (Yates and Grundy, (56)) for fixed size n will include both the inclusion probability of unit i and the joint inclusion probability of units i and j ($i \neq j$). The minimization of this expression with respect to the inclusion probabilities requires knowledge of the parameter $(y_1, \ldots, y_N)$.

C. Uniform Admissibility

Joshi (30) broadened the idea of admissibility to include the sample design in defining what he called "uniform admissibility".

<u>Def</u>. An estimator e_0 and sampling design p_0 are uniformly admissible for the population total T(y) if there does not exist any other estimator e_1 and sampling design p_1 such that

i) Expected sample size for $p_1 \leq$ Expected sample size for p_0

ii) M.S.E. $(e_1, p_1) \leq$ M.S.E. (e_0, p_0) and strict inequality holds either in i) or for at least one $y \varepsilon R^N$ in ii).

Joshi's justification of uniform admissibility as a criterion is the following:

i) Admissibility alone compares estimators with respect to a fixed design, p_0. For a given p_0, an estimator e_0 may be admissible but the estimator e_1 may not. However, there may be another sampling design p_1 for which e_1 is admissible and M.S.E. $(e_1,p_1) \leq$ M.S.E. (e_0,p_0).

ii) In practice, considerations of cost and time are important. If there exists a p_1 such that M.S.E. $(e_1,p_1) \leq$ M.S.E. (e_0,p_0) and the expected sample size $(p_1) \leq$ expected sample size of (p_0), then it would be wise to use (e_1,p_1) in the survey design.

Joshi showed that the estimator

$$e^*(s,y) = \frac{N}{n(s)} \sum_{i\varepsilon s} y_i$$

where n(s) is the number of distinct units in the sample, and any fixed sample size design p* are uniformly admissible for the population total $N\bar{Y}_N$. Obviously, a large class of sampling designs satisfy the conditions in Joshi's results.

D. Expected Variance

Cochran (4) compared the variance of the sample mean for stratified random sampling, one unit per stratum, with the variance of the mean for systematic sampling (sys.). He made the assumption that the elements were serially correlated, the correlation between two elements being positive and a monotone decreasing function of the distance apart of the elements. He then pointed out that to assume the correlations to be strictly monotone for an actual finite population would not be realistic. Although the correlogram may be downward sloping, individual fluctuations about the trend may prevent it being strictly monotone. Hence, he regarded the finite population as being a sample from an infinite population in which the correlations are monotone. Comparison of the two survey designs was on the basis of the expectation of their respective variances over all finite populations that could be drawn from the infinite one.

Specifically, Cochran assumed the infinite population to have the parameters

$$\mathcal{E}(y_i) = \mu \quad \forall_i$$

$$\mathcal{C}ov(y_i, y_j) = \sigma^2 \quad i = j$$

$$= \rho_u \sigma^2 \quad i \neq j \quad \text{where } j - i = u \text{ and}$$

$$\rho_u \geq \rho_v \geq 0 \; , \quad u < v \; .$$

The utilization of the expected variance of the estimator over all finite populations as a basis for selecting the sample design and estimator is now commonly called the expected variance criterion. We denote the expected variance by $\mathcal{E}V$. The sample design $p \in \underline{P}$ and estimator $a \in A$ with smallest $\mathcal{E}V$ are then said to be best with respect to the infinite population. Cochran's main result is that for all infinite populations in which

$$\rho_i \geq \rho_{i+1} \geq 0 \quad i = 1, 2, \ldots, N-1 \quad \text{and}$$

$$\rho_{i-1} + \rho_{i+1} - 2\rho_i \geq 0 \quad i = 1, \ldots, N-2$$

$$\mathcal{E}V(\text{sys.}) \leq \mathcal{E}V(\text{strat.}) \leq \mathcal{E}V \quad \text{(simple random sampling).}$$

Godambe (16) attempted to justify the expected variance as a criterion for the selection of estimators. His reasons were the following:

i) "On the basis of past experience regarding several factors which influence the value of the variate y under study, or because of the knowledge of the distribution of one or more correlated values, the statistician often may have certain expectations of y associated with different individuals in the population."

ii) "These expectations are a priori expectations in the sense that they exist before drawing and admit simple interpretations in terms of a priori probabilities."

Godambe formalized the a priori information as follows: $\mathcal{E}(y_\lambda)$ denoted the a priori expectations, $\mathcal{V}(y_\lambda)$ the apriori variances and it was assumed that the a priori covariances of $y_\lambda, y_{\lambda'}$, $\lambda \neq \lambda'$ $(\mathcal{C}ov(y_\lambda, y_{\lambda'}))$ were zero. Further, it was assumed that the $\mathcal{E}(y_\lambda)$ did not change after drawing one or more units from the population and observing the values of y associated with them. Godambe also stated that if the estimator depended on $\mathcal{V}(y_\lambda)$ it would be of little practical use since $\mathcal{V}(y_\lambda)$ is almost never known.

Godambe attempted to minimize the $\mathcal{E}V(e,p)$ with respect to the subclass of estimators

$$T_5 = \left\{ t_5 = \sum_{i \varepsilon s} \beta_{is} y_i \right\}$$

where β_{is} is the coefficient of the i^{th} unit in the s^{th} sample and over all possible $p \varepsilon \underline{P}$. The notation T_5 appears in Koop (31). Minimization of $\mathcal{E}V(e,p)$ resulted in difficulty and so Godambe minimized $\mathcal{E}V(e,p)$ with respect to a subclass of designs D^n where $d \varepsilon D^n$ satisfied,

i) d is a fixed sample size design, units selected without replacement

ii) the inclusion probability of unit u_λ ,

$$\pi_\lambda = \frac{n\,\mathcal{E}(y_\lambda)}{\sum_{\lambda=1}^{N} \mathcal{E}(y_\lambda)} \quad , \quad \mathcal{E}(y_\lambda) > 0 \ \forall\lambda \ .$$

Godambe then showed that

$$\mathcal{E}\, V(t_5, d) \geq \sum_{\lambda=1}^{N} \frac{\mathcal{V}(y_\lambda)}{\pi_\lambda} - \sum_{\lambda=1}^{N} \mathcal{V}(y_\lambda) \quad .$$

Further, he showed that the H.T. estimator satisfied this lower bound. Hence it was best. Godambe also showed that if

i) $\mathcal{V}(y_\lambda) = c\,\mathcal{E}(y_\lambda)^2$ where c is a positive constant

ii) $\underline{P}^* = \{p' : p'$ is a fixed sample size design selecting units without replacement$\}$,

then

$$\mathcal{E}\, V(\text{H.T. estimator}, p \varepsilon D^n) \leq \mathcal{E}\, V(t_5, p' \varepsilon \underline{P}^*) \quad .$$

Hájek (20) also restricted A to unbiased estimators in T_5 and attempted to find the "best" estimator and sampling design. Quoting Hájek, "the notion of optimum strategy is useless when only one particular sequence $y_1, \ldots, y_N$ is considered, because the problem disappears if we know it, and has no solution if we don't." Hájek also used the expected variance approach, i.e. he supposed there existed a certain probability distribution with mean vector $\underline{\mu}$ and variance-covariance matrix $\underline{\Sigma}$ in the space of sequences $y_1, \ldots, y_N$ and searched for an unbiased estimator in T_5 and $p \varepsilon \underline{P}$ which would minimize the $\mathcal{E}\, V(t_5, p)$. Hájek justified this approach on the grounds that

i) In most cases, one has some knowledge of conditions producing values $y_1, \ldots, y_N$ and can express it in the form $\underline{\mu}$ and $\underline{\Sigma}$.

ii) The assumption of a probability distribution with parameters $\underline{\mu}$ and $\underline{\Sigma}$ influence the sampling-estimating strategy but do not influence the validity of the estimated sampling errors. Any mistake in $\underline{\mu}$ and $\underline{\Sigma}$ will only cause the sampling errors to be on the average greater than if they were correct.

When

$$\underline{\mu} = \begin{bmatrix} \xi(y_1) \\ \cdot \\ \cdot \\ \cdot \\ \xi(y_N) \end{bmatrix}$$

and

$$\underline{\Sigma} = \begin{bmatrix} \gamma(y_1) & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \gamma(y_N) \end{bmatrix}$$

and a cost function of the form

$$C = \sum_{i=1}^{N} c_i \pi_i$$

(π_i : inclusion probability of the i^{th} unit) is assumed, Hájek showed that any survey design such that

i) π_i proportional to $\sqrt{\dfrac{\gamma(y_i)}{c_i}}$

ii) $\beta_{is} = \frac{1}{\pi_i}$

iii) $\sum_{i\varepsilon s} \beta_{is} \mathscr{E}(y_i) = \sum_{i=1}^{N} \mathscr{E}(y_i)$

for all s such that $p(s) > 0$ is the best one in the sense that it minimizes the $\mathscr{E}V(t_5,p)$ for a fixed cost.

Such a design may not exist but if $\mathscr{E}(y_i) = \mu$, $\mathscr{V}(y_i) = \sigma^2$, $c_i = c$, $\forall_i$, he showed that simple random sampling without replacement and the sample mean satisfied conditions i), ii) and iii).

Hájek (20) also expanded the $\mathscr{E}V(t_5,p)$ in the form

$$\mathscr{E}V(t_5,p) = \sum_s\left(\sum_{i\varepsilon s} \beta_{is}\mathscr{E}(y_i) - \sum_{i=1}^{N} \mathscr{E}_i\right)^2 p(s)$$

$$+ \sum_{i=1}^{N} \mathscr{V}(y_i)\left\{\sum_{s\ni i} (\beta_{is}-1)^2 p(s) + (1-\pi_i)\right\}$$

where

$$\sum_{s\ni i}$$

denotes summation over all samples containing unit i. This illustrated that the minimization of $\mathscr{E}V(t_5,p)$ involved the examination of two components. The first component contained the a priori means and the second, the a priori variances.

Under a somewhat more general a priori set-up, i.e.,

$\mathscr{E}(y_i) = \mathscr{E}_i > 0 \ \forall_i$

$$\mathscr{C}ov(y_i, y_j) = K\xi_i^{\,2} \quad i = j \quad K > 0$$
$$= \xi_i \xi_j \, R(|i-j|) \quad i \neq j$$

where $R(|i-j|)$ is a convex function, Hájek showed that unequal probability systematic sampling with $\pi_i \propto \xi_i$ and the H.T. estimator is the optimum survey design. Note that Hájek assumed that the a priori variances would be known and be important in the construction of the estimator and sampling design. Godambe (16) on the other hand felt that the a priori variances would rarely be available to the survey designer.

E. Hyper-admissibility

Hanurav (25) introduced the concept of Hyper-admissibility which is weaker than uniform minimum variance in that a uniform minimum variance unbiased estimator is also hyper-admissible. On the other hand an admissible estimator is not necessarily hyper-admissible.

The parameter $(y_1, \ldots, y_N)$ can be envisioned as a point in R^N space. Let e be an unbiased estimator of the population total T. An unbiased estimator for any linear combination,

$$\sum_{i=1}^{N} \ell_i y_i \quad ,$$

of the y_i's can be obtained by replacing y_i in e by $\ell_i y_i$.

For example if ℓ_i = 0 or 1 for all i, let K ℓ_i's be 1. Then an estimator e is hyper-admissible if it is admissible in R^N and if the estimator e* created by substituting $\ell_i y_i$ for y_i is admissible in R^K. Thus hyper-admissibility requires that e be admissible not only in the whole of R^N but also in each of its principal hyperplanes. Formally, Hanurav (22) stated, "In the class of unbiased estimators of T, e in T is hyper-admissible if it is admissible when the parameter y is restricted t, the interior of any principle hyper surface R^{N-}."

Hanurav justified hyper-admissibility by arguing that in the estimation of linear parametric functions

$$\sum_{i=1}^{N} \ell_i y_i \quad ,$$

in which some of the ℓ_i are known to be zero, the parameter space of relevance should no longer be R^N but a hypersurface of R^N and the estimator should be admissible in this hypersurface. If another linear parametric function becomes of importance at a later date, a new type of estimator may be required if the original estimator is not hyper-admissible.

F. Superpopulation

The concept of superpopulation is present in any discussion of expected variance. Such an assumption is a part of the prior information I. In most of the literature knowledge of the distribution of the superpopulation is limited to the mean, variance and covariance. Rosenzweig (49) assumed knowledge of the family of underlying distributions. He was interested in the presence of large units in the sample. He assumed all elements beyond some value y_0 were selected from a truncated Weibull distribution. It was desired to estimate the population mean utilizing the knowledge of the form of the underlying distribution. The observations larger than x_0 were tested for skewness. If the null hypothesis (underlying distribution is exponential) was not rejected, the simple mean was used to estimate the mean of these elements; if it was rejected, a linear estimator based on order statistics was used. The overall estimator of the population mean was based on a weighted average of the means of the two groups.

When information on the type of distribution is available, as opposed to knowledge of the moments only, a different approach is utilized. Whereas Godambe (16) and Hájek (20) used the moments in their choice of $p \varepsilon \underline{P}$ and $a \varepsilon A$, Rosenzweig used his knowledge of the distribution to construct preliminary tests of the distribution parameters to obtain an estimator.

Another superpopulation result is that of Blackwell and Girschick (3). The authors discuss the concept of a statistical game involving a finite population. Given a sample space (Y, Ω, P), a fixed-sample size experiment is performed and a value of a random variable $y = (y_1, \ldots, y_N)$ is obtained. The statistician has to select one out of a class A of possible actions (by actions is meant an estimator $a \varepsilon A$). He incurs a loss L dependent on y and $a \varepsilon A$. The problem is to determine the optimum method of selecting a sample of size n from the N elements when the cost of sampling is independent of the units i. Further structure is placed on the problem, namely;

i) A sample space (Y, Ω, P) where $y = (y_1, \ldots, y_N) \varepsilon Y$ and $Y = \Omega$; $p(y|\omega) = 1$ if $y = \omega$

$= 0$ otherwise

ii) arbitrary space of actions

iii) a set S of all subsets of the integers 1 to N of size n

iv) a decision function space $S \times D$ with elements (s,d) where d depends only on the coordinates specified by s and $d(y) \varepsilon A \ \forall y$.

v) A class of random procedures $\Phi = \{ \phi \ni \phi$ is defined on $S \times A \times Y$, nonnegative and such that

a) $\sum_{s \varepsilon S} \sum_{a \varepsilon A} \phi(s, a|y) = 1 \quad \forall y$

b) $\sum_{a \varepsilon A} \phi(s, a|y)$ is independent of y $\forall \ s \varepsilon S$

c) for each $s \varepsilon S$, $\phi(s, a|y)$ depends only on the coordinates of $y \}$.

vi) A group G of elements g^T defined on S x A x Y ∋

$$g^T(y,\omega,\bar{a}) = (g_Y{}^T(y),\ g_\Omega{}^T(\omega),\ g_{\bar{A}}^T(\bar{a}))$$

where T is a permutation of the integers 1,2,...,N and $\bar{A} = S \times A$.

vii) A loss function L defined on Ω x A and does not depend on s and constant ∀ permutations of the coordinates of $\omega = y$.

Invoking the Principles of Invariance and Sufficiency they show that the strategy of simple random sampling is the preferred action. The authors do not consider the estimation of a function of $y = (y_1,\ldots,y_N)$.

G. Minimax Criterion

Aggarwal (1) considered the estimation of the mean of a finite population. He first assumed that

i)
$$\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix} \Big| \omega \sim \text{Normal}_N \left(\begin{bmatrix} \mu_\omega \\ \cdot \\ \cdot \\ \cdot \\ \mu_\omega \end{bmatrix}, \begin{bmatrix} \sigma^2\,\frac{N-1}{N} & & -\frac{\sigma^2}{N} \\ & \ddots & \\ & & \sigma^2\,(\frac{N-1}{N}) \end{bmatrix} \right)$$

ii)
$$E_\omega \sum_{i=1}^{N} (y_i-\mu_\omega)^2 = \int \cdots \int_{R^N} (y'y-N\mu^2_\omega)\,d_\omega(y) \leq (N-1)\sigma^2$$

where

iii) The sample space is (Y,Ω,P) where Y is in R^N,

Ω is the set of all distributions ω on hyperplanes in R^N of the form $y_1 + y_2 + \ldots + y_N = N\mu_\omega$ and subject to ii) above.

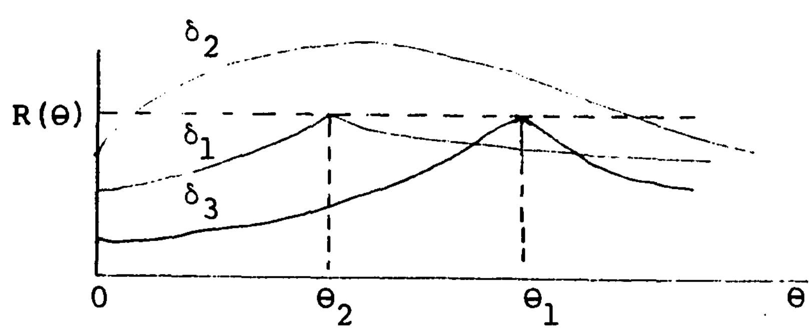
iv) The action space A is the real line R^1

v) The loss function L, defined on $(\Omega \times A)$ is given by $L(\omega, a) = (a - \mu_\omega)^2$.

He reduced his search of $a \varepsilon A$ to those estimators δ which were minimax, i.e. which minimized

$$\sup_{\omega \varepsilon \Omega} E_\omega L(\omega, a)$$

where E_ω denotes expectation with respect to ω. Referring to Blackwell and Girschick's result Aggarwal assumed SRS without replacement. He showed that if nature's strategy was to pick μ_ω from a $N(0, \theta^2)$, the sample mean is a minimax estimator. Further, he showed that the usual stratified mean is minimax. Minimax estimators however are not unique. A simple graphical example is shown below. Suppose there exists



only three possible estimators δ_1, δ_2, δ_3 of a parameter θ. Denote $R_i(\theta)$ the risk associated with estimator δ_i.

Clearly if $R_1(\Theta_2) = R_3(\Theta_1)$, δ_1 and δ_3 are minimax in this restricted class of estimators. However, if it is suspected that Θ is "close to zero" then δ_3 is to be preferred to δ_1 even though δ_1 is also minimax.

III. THE ANTICIPATED VARIANCE

Specifying the prior information in the form of knowledge of the superpopulation from which the finite population was selected is one possible way in which information available to the survey designer may be formulated. In some cases, information is limited to the first order moments and in other cases to the form of the superpopulation. Other types of prior information involve the specification of a linear relationship between y and a known variable x or the specification of a monotone trend in the expected value of the y's. In this study, we shall assume that the survey designer has some knowledge of the vector of expectations ξ and the covariance matrix γ given by

$$\xi = \begin{bmatrix} \xi_1 \\ \cdot \\ \cdot \\ \cdot \\ \xi_N \end{bmatrix}$$

and

$$\gamma = \begin{bmatrix} \gamma_{11}, \zeta_{12} & \cdots & \zeta_{1N} \\ \cdot & & \\ & \cdot & \\ \zeta_{N1} \zeta_{N2} & \cdot & \gamma_{NN} \end{bmatrix} \qquad \text{(III.1)}$$

where $|\gamma| > 0$.

This specification of the prior information is similar to that employed by Cochran (4), Godambe (16) and Hájek (20) when ξ and γ are interpreted as the first two moments of a superpopulation. Occasionally the prior information may specify the vector of values as a function of a concomittant variable x_i (e.g. βx_i with β unknown). The designer is expected to know the ratios γ_{ii}/γ_{jj} $i \neq j$, prior to sampling and , multiplying γ_{ii} by an unknown constant $\sigma^2 > 0$ will not alter the ratio, γ_{ii}/γ_{jj}. The importance of the knowledge of γ_{ii}/γ_{jj} will become evident later.

To keep the estimator and design from being entirely model dependent, we shall restrict the estimator to be unbiased and (or) consistent for $\bar{Y}_N$ with respect to the sampling design irrespective of any model assumptions.

Since the survey designer has only prior information at his disposal in designing the survey, the measure of "goodness" for his survey design must be a function of his prior information. As a measure of "goodness" we define the Anticipated Variance as follows:

Def. The Anticipated Variance (A-V) of estimator a of the finite population mean under the sampling design p is the variance of the expression $a-\bar{Y}_N$ treating the imperfect prior information as if it were true.

This definition assumed that the prior information is expressible in a probabilistic manner. The notation A-V($a-\bar{Y}_N$) can be used to denote the anticipated variance of a as an estimator of $\bar{Y}_N$. Where no confusion will result we shall abbreviate this notation to A-V(a).

If the entities in III.1 are interpreted as moments of a superpopulation from which the finite population has been selected the A-V(a) is the variance of the expression $a-\bar{Y}_N$ over all samples n and all finite collections of N random variables from a superpopulation with the specified moments. The situation is as follows:

Given the vector of values ξ, "nature" draws a random variable $(e_1,\ldots,e_N)$ from the given superpopulation characterized by a zero mean vector and covariance matrix $\mathcal{V}$. The values $y_i = \xi_i + e_i$, $i = 1,2,\ldots,N$ comprise the finite population of size N. The survey designer wishes to select a sample of size n, say, $(y_1,\ldots,y_n)$ from $(y_1,y_2,\ldots,y_N)$ to estimate $\bar{Y}_N$, the finite population mean. Under the superpopulation model, $\mathcal{E}(y_i|i) = \xi_i$ where $\mathcal{E}(y_i|i)$ is to be interpreted as 'the average value of y_i over all finite populations of size N.' The notation '$|i$' is redundant since the integer i is used to designate the population unit of interest. The notation is not meant to convey a conditional expectation in the probabilistic sense. A similar interpretation is given to

$\mathcal{V}(y_i|i)$, (variance of y_i) and $\mathcal{C}ov(y_i,y_j)|i,j)$, (covariance of y_i and y_j).

There are two sources of randomness under the model. One source of randomness is associated with selection of the finite universe from the superpopulation. The other source of randomness is associated with the sampling design p. The selection of the e_i is assumed to be independent of the sampling design. This is a reasonable assumption since the sampling design is a function of the units (1,2,...,N) and not a function of the character y associated with the units.

The sample of size n, $(y_1,\ldots y_n)$, from $y_1,\ldots,y_N)$ can also be considered a sample from the superpopulation. Although we sample from $(y_1,\ldots,y_N)$, we only observe $(y_1,\ldots,y_n)$. Thus, given a sample of size n, $(y_1,\ldots,y_n)$ we may fix the indices (1,2,...,n) and average over all finite populations of size N containing $(y_1,\ldots,y_n)$. We then obtain the expected value, under the model, of the estimator conditional on the sample containing the elements indexed by $i = 1,2,\ldots n$.

When the model is $y_i = \beta x_i + e_i$ with $(e_1,\ldots e_N)$ distributed with mean vector 0 and covariance matrix $\mathcal{V}$, we may assume that each unit i in the population has two related characteristics y_i and x_i and selection of a sample of size n will result in the observation of n tuples $(y_1,x_1),\ldots,$ (y_n,x_n). In this case $\mathcal{E}(y_i|i)$ will be interpreted as meaning $\mathcal{E}(y_i|x_i)=\beta x_i$. Furthermore, to simplify the notation,

$\mathscr{E}(a-\bar{Y}_N|x_1,\ldots,x_n)$ will be written as $\mathscr{E}(a-\bar{Y}_N|s)$. Similarly, $\mathscr{V}(a-\bar{Y}_N|x_1,\ldots,x_n)$ will be written $\mathscr{V}(a-\bar{Y}_N|s)$.

For example, if in III.1, $\mathscr{C}_{ij} = 0$ for $i \neq j$, and

$$a = \sum_{i=1}^{n} \beta_{is} y_i$$

is unbiased for $\bar{Y}_N$ under p, then

$$\mathscr{E}(a-\bar{Y}_N|s) = \sum_{i=1}^{n} \beta_{is}\mathscr{E}_i - \bar{\mathscr{E}}_N$$

and

$$A\text{-}V(a) = \sum_s\Big(\sum_{i=1}^{n} \beta_{is}\mathscr{E}_i - \bar{\mathscr{E}}_N\Big)^2 P(s) + \sum_{i=1}^{N} \mathscr{V}_i\Big[\sum_{s\ni i}(\beta_{is}-1)^2 P(s)$$

$$+ (1-\pi_i)\Big] \quad .$$

The survey designer may not believe in a superpopulation model. However, when asked to "guess" at a value for unit i, he is often able to give some value $\mathscr{E}_i$ which he feels will approximate y_i. The process through which the designer arrives at $\mathscr{E}_i$ is his assessment of the situation based on his "experience". For example, if he is estimating income, $\mathscr{E}_i$ may represent last year's income of unit i increased by a percentage cost of living factor. Needless to say, $\mathscr{E}_i$ is rarely the true value of y for unit i. In this non-superpopulation model situation, $\mathscr{E}_i$ represents the designer's anticipation for y_i.

We assume that ξ_i represents an "average" of all "feelings" the designer has concerning the value y_i. Such an "averaging" process is of a subjective nature, the anticipated value ξ_i being a function of the possible values ξ_j^i (his "feelings") and their degrees of "likelihood" P_j (those ξ_j^i's that he feels more likely to be y_i will have larger P_j's), $j = 1, \ldots, M^i$; M^i finite. No two designers can be expected to arrive at the same values for ξ_j^i and P_j, nor would it be realistic to assume that every designer obtains his anticipated values by specifying ξ_j^i and P_j, $j = 1, \ldots, M^i$ for all i. The designer may often state the value ξ_i without specifying the ξ_j^i and P_j. However, he is assumed to be performing, subconsciously or "within himself", an averaging process in order to obtain ξ_i.

In addition to specifying ξ_i, the designer may also have a feeling or a measure of "confidence" or "accuracy" for the ξ_i. It is natural to draw a parallel between this measure of "confidence" and a variance. For example, he may be able to specify a range about ξ_i within which he expects the value y_i to fall, or he may be able to give ratios of the ranges for any two units i and j. Hence, we assume the measure is quantifiable and denote it by γ_i. When $\gamma_i > \gamma_j$, $i \neq j$, the designer feels that on the average y_j is closer to ξ_j than y_i is to ξ_i. In III.1 positive ℓ_{ij} will occur if, for example, the survey designer feels, "If I overestimate on unit i, I expect to also overestimate on unit j."

It is reasonable to assume that given the units (1,2,...,n) the designer's anticipated value for the estimator

$$a = \sum_{i=1}^{n} \beta_{is} y_i$$

and the population mean, $\bar{Y}_N$, will be

$$\sum_{i=1}^{n} \beta_{is} \xi_i$$

and $\bar{\xi}_N$, respectively. Similarly, given the units (1,2,...,n), we assume he will be able to specify ranges of y_i about ξ_i which behave as standard errors and enable one to calculate the $\mathcal{V}(a - \bar{Y}_N | s)$ obtaining a result similar to that obtained under the superpopulation model.

The values of $\xi(a - \bar{Y}_N | s)$ and $\mathcal{V}(a - \bar{Y}_N | s)$ being formal functions of his original beliefs are purely personal. We say that $\xi(a - \bar{Y}_N | s)$ and $\mathcal{V}(a - \bar{Y}_N | s)$ are the designer's anticipated value and anticipated variance, respectively, for $a - \bar{Y}_N$ when he is given the sample (1,2,...,n).

If for any unit i, $\mathcal{V}_i = 0$, this implies that the designer is certain that $\xi_i = y_i$ and hence he may choose to sample from the remaining units, omitting unit i. In the limiting case, with all $\mathcal{V}_i$ zero, the designer need not sample since $\mathcal{V}(a - \bar{Y}_N | s) = 0$ for all samples.

Under the superpopulation model, the anticipated variance of an estimator of $\bar{Y}_N$ is the variance of the expression $a-\bar{Y}_N$ over all samples n and all finite collections of N random variables from the superpopulation with the specified moments. In the superpopulation model, $\mathcal{E}$ represents the mean vector of an $N \times 1$ vector valued random variable.

When the superpopulation model is not assumed, the anticipated variance of an estimator of $\bar{Y}_N$ is the variance of the expression $a-\bar{Y}_N$ over all samples n and all personalistic evaluations of the N units denoted by III.1. In this case, the averaging process is entirely a function of the designer's "feelings". We maintain that in this case a probabilistic interpretation of III.1 is still possible. $\mathcal{E}$ and $\mathcal{V}$ now representing the designer's personalistic or subjective evaluation (expectation) operators.

It will often be convenient to express the A-V as

$$A\text{-}V(a) = V[\mathcal{E}(a-\bar{Y}_N|s) + E[\mathcal{V}(a-\bar{Y}_N|s) \qquad \text{(III.2)}$$

where V and E are the variance and mean operators with respect to the sampling design $p \varepsilon P$ and $\mathcal{E}$ and $\mathcal{V}$ are the operators with respect to the designer's subjective model assumption.

The minimization of the A-V is a way to utilize the available prior information to select the estimator and sampling design that minimize the variance under the subjectively hypothesized model. With either a superpopulation or personalistic interpretation of the prior information, utilization of the A-V seems reasonable.

The following is an example of a situation when the type of prior values in III.1 may occur.

We are interested in estimating the mean income of the membership of a large professional society, e.g. ASA. We shall define membership as all those persons paying dues in a specified time period.

Societies with large memberships are scattered throughout many regions of the country. In each region, the society has a chapter and a secretary who conducts the business of the regional group. The secretary is personally acquainted with many members of the chapter and ordinarily will have an idea about their incomes.

For purposes of exposition, assume the society is partitioned into 10 chapters. To estimate the mean income of the membership in chapter A we contact the secretary of the chapter. The secretary provides us with a list of the members in chapter A. We then ask him what he "feels" is the income (within a hundred dollars) of each of the names on the list. Some of the members are close friends, and he can state their incomes without qualification. Others are merely passing acquaintances but knowing their positions and the companies they work for, he can give a guess as to their incomes with a qualification that he may be in error, "give or take \$1500.00." In the same vein, he may know that

John and Bob work in the same department and perform the same duties and that Bob who earns $12,000.00 is three years older than John and has worked there three years longer. Hence a positive correlation is evident.

We can then transform the feelings of the secretary into the form of III.1 and proceed to design the survey using this prior information. Since ten different secretaries are involved, we might stratify the membership into 10 strata and sample independently within each strata.

In the next chapter, several types of models will be investigated. The A-V criterion will be used and estimators and sampling designs will be obtained to minimize the A-V.

IV. SIMPLE AND RATIO MODELS

The choice of a survey design depends on the type of prior information available to the survey designer. In this chapter, we shall specify particular models, albeit simple ones, for the form of the prior information in III.1. We shall consider fixed sample size designs only. That is to say, we shall consider sample designs P that assign probability greater than or equal to zero to samples s that are of size n and zero probability to all other samples. In addition we shall assume the presence of a complete frame and complete response of units drawn for the sample.

Suitable reductions of the class of estimators and designs will be made when it becomes evident that a general approach is intractable. The A-V criterion will be used to determine the best survey design under the hypothesized model.

We give definitions of terms to be used throughout the remainder of the dissertation.

Def. A sample design p is said to be of fixed size n if all samples s such that $p(s) > 0$ contain n units, not necessarily distinct.

Def. A sample design p is said to be of variable size if it is not of fixed size n.

Def. A sample design p is said to be with replacement of size n if

i) it is of fixed size n and

ii) given probabilities p_i, $i=1,\dots,N$ and

$$\sum_{i=1}^{N} p_i = 1 \quad ,$$

selection of the n units is performed by randomly selecting unit i with probability p_i at every draw.

Selection of the units is performed independently from draw to draw. Thus the size can be thought of as the number of draws.

Def. Given a sampling design p, an estimator e is said to be unbiased admissible if

i) e is unbiased

ii) there does not exist another unbiased estimate e' such that $V(e') \leq V(e)$ for any $(y_1,\dots,y_N) \varepsilon R^N$ with strict inequality holding for at least one $(y_1,\dots,y_N)$.

The symbol V denotes the variance operator. This definition differs from the usual definition of admissibility in that it restricts consideration to unbiased estimators.

Def. An estimator e of $\bar{Y}_N$ is said to be unbiased under the model if given a sample s, the expectation of $e-\bar{Y}_N$ over all possible infinite populations with the specifications of the model is equal to zero. In the previously introduced symbolism, e is unbiased under the model if $\mathcal{E}(e-\bar{Y}_N|s) = 0$.

If we use the term unbiased it may be presumed to mean unbiased with respect to the sampling design p, i.e.,

$$\sum_s e(s)p(s) = \bar{Y}_N.$$

A. Model I

The first model to be considered is

$$\mathcal{E}(y_i|i) = \mu$$

$$\mathcal{C}ov(y_i,y_j|i,j) = \rho\sigma^2 \quad i \neq j \quad -\frac{1}{N-1} \leq \rho < 1$$

$$= \sigma^2 \quad i = j$$

where μ, ρ, and σ^2 are unknown. We shall call this model the 'constant case' model, for obvious reasons. The 'constant case' model may arise in situations where the survey designer is unfamiliar with the population of interest. Such will be the case when no previous survey has been conducted to study the population.

We shall first state and prove results for several sub-problems of interest. These sub-problems are characterized by limitations on the class of estimators or designs or by further restriction of the 'constant case' model. We state the first sub-problem as Theorem A.1.

Theorem A.1:

Given that

i) we sample w/o replacement of size n

ii) π_i denotes the inclusion probability of unit i.

In order to have $\bar{y}_n$ unbiased for $\bar{Y}_N$, we must have $\pi_i = \frac{n}{N}$.
For any design satisfying i),

$$A\text{-}V(\bar{y}_n) = \sigma^2\left(\frac{1}{n}-\frac{1}{N}\right) + \rho\sigma^2\left(\frac{n-1}{n}-\frac{N-1}{N}\right) \quad .$$

Proof:

The proof is given in Appendix 1.

Hence simple random sampling without replacement (denoted SRS w/o) of size n meets the specifications of the theorem. Equal probability systematic sampling of size n will also suffice (when $nL = N$ and L is integral). Evidently, in this case, the A-V criterion alone is insufficient to completely determine the best sampling design.

In Theorem A.1 we fixed the estimator and found a suitable sampling design. In sub-problem A.2 we do the converse, i.e., we fix the sampling design and find the best estimator in a certain class of estimators.

Theorem A.2:

Given that Model I holds, let T_7 denote the class of estimators

$$t_7 = \Sigma c_{s_{i_t}} y_i$$

where

$$c_{s_{i_t}}$$

is the coefficient of element u_i whenever it appears at the t^{th} draw in the s^{th} sample. Assume also that

i) we wish t_7 to be unbiased for $\bar{Y}_N$ and

ii) we wish to sample SRS w/o replacement of size n.

Then, the sample mean, $\bar{y}_n$, satisfies i) and possesses the smallest A-V of all estimators in T_7.

Proof:

The details of the proof are presented in Appendix 2.

In Theorems A.1 and A.2, either the estimator was fixed and the best sampling design derived or the sampling design was fixed and the best estimator of a class of estimators derived. We now state a theorem simultaneously yielding the best sampling design and estimator.

Theorem A.3:

Given Model I. Also, assume that

i) $(y_1, \ldots, y_N)$ are independent with respect to some underlying superpopulation or personalistic distribution and hence $\rho = 0$,

ii) we wish to use an unbiased estimator of $\bar{Y}_N$

iii) we wish to use a sample design p such that

$$\sum_{i=1}^{N} \pi_i = n \ , \quad \text{and}$$

iv) we wish to minimize the A-V of the estimator.

Then, the estimator $\bar{y}_n$ together with SRS w/o replacement of size n satisfy ii) and iii) respectively, and jointly satisfy iv).

Proof:

Godambe and Joshi (17) have shown that in the class of designs admitting unbiased estimators of $\bar{Y}_N$, any unbiased estimator will have A-V greater than or equal to

$$\frac{1}{N^2} \sum_{i=1}^{N} \frac{\sigma^2}{\pi_i} - \frac{\sigma^2}{N} \ .$$

It can be easily verified that the A-V($\bar{y}_n$) under SRS w/o replacement of size n satisfies the lowest possible bound under the conditions stated in the hypothesis. Q.E.D.

It should be noted that Godambe and Joshi's (17) result need not necessarily be restricted to fixed sample size designs. Also note that our sample design is optimum over all sample designs whose expected number of distinct units is n and not necessarily over fixed sample size designs only. However, since variable sample size designs are often undesirable from the administrative point of view, we shall restrict ourselves to fixed sample size designs.

In the next theorem we attain generality in the type of model information but simultaneously introduce a restriction in the class of estimators.

Theorem A.4:

Given Model I with $\rho \neq 0$. Assume also that we wish

i) to use an unbiased estimator t_7 in T_7,

ii) to use a fixed sample size design of size n, and

iii) we wish to minimize the A-V(t_7).

Then, $\bar{y}_n$ and SRS w/o replacement of size n satisfy i) and ii) respectively, and jointly satisfy iii).

Proof:

The proof is given in Appendix 3.

To summarize the results in section A, under the 'constant case' model, $\bar{y}_n$ together with SRS w/o replacement of size n are jointly best in a wide class of designs.

B. Model II

The second model is

$$\mathcal{E}(y_i|x_i) = x_i \quad \forall_i$$

$$\mathcal{C}ov(y_i,y_j|x_i,x_j) = \gamma_i \quad i = j$$

$$= 0 \quad i \neq j \quad .$$

We shall assume that $\bar{X}_N$ and γ_i/γ_j for all $i \neq j$ are known to the survey designer prior to sampling. A sample of size n will consist of n tuples (y,x) where y is the characteristic in question and x is a concomittant variable which may or may not be known prior to sampling. Given the above information we have,

Theorem B.1:

Given

i) Model II with the added assumption that $(y_1,\ldots,y_N)$ are independent with respect to some underlying personalistic distribution or superpopulation.

ii) the estimator e must be unbiased,

iii) the nonreplacement sampling design must be such that

$$\sum_{i=1}^{N} \pi_i = n \qquad \text{and}$$

iv) $0 < \dfrac{n\sqrt{\gamma_i}}{\sum_{i=1}^{N}\sqrt{\gamma_i}} < 1$ for all i.

Then, the estimator

$$d_1 = \sum_{i\varepsilon s} \frac{y_i}{N\pi_i} + \bar{X}_N - \sum_{i\varepsilon s} \frac{x_i}{N\pi_i}$$

and the nonreplacement sampling design p such that

$$\pi_i = \frac{n\sqrt{\gamma_i}}{\sum_{i=1}^{N}\sqrt{\gamma_i}}$$

for all i satisfy ii) and iii) respectively, and jointly minimize the A-V(e). If for some i, iv) does not hold, we may include these elements in the sample with certainty and complete the sample by selecting units from the reduced finite population.

Proof:

It is easy to see that d_1 is unbiased for $\bar{Y}_N$ if we sample with inclusion probability π_i. The

$$A\text{-}V(d_1) = E\left[\sum_{i\varepsilon s} \frac{\mathcal{V}_i}{N^2\pi_i^2} + \sum_{i=1}^{N} \frac{\mathcal{V}_i}{N^2} - 2 \sum_{i\varepsilon s} \frac{\mathcal{V}_i}{N^2\pi_i}\right]$$

$$= \sum_{i=1}^{N} \frac{\mathcal{V}_i}{N^2\pi_i} - \sum_{i=1}^{N} \frac{\mathcal{V}_i}{N^2} .$$

Since the second term is fixed, we minimize the first term with respect to iii).

Let H be the Lagrangian

$$H = \frac{1}{N^2} \sum^{N} \frac{\mathcal{V}_i}{\pi_i} + \lambda\left(\sum_{i=1}^{N} \pi_i - n\right)$$

where λ is the Lagrangian multiplier. Now

$$\frac{\partial H}{\partial \pi_i} = - \frac{\mathcal{V}_i}{N^2\pi_i^2} + \lambda = 0 \quad i = 1, \ldots, N . \tag{IV.1}$$

Multiplying the above N equations by π_i and summing over N, we have

$$- \frac{1}{N^2} \sum_{i=1}^{N} \frac{\mathcal{V}_i}{\pi_i} + n\lambda = 0 \tag{IV.2}$$

and hence

$$\lambda = \frac{1}{N^2 n} \sum_{i=1}^{N} \frac{\gamma_i}{\pi_i} \quad .$$

Substituting into IV.1, we have

$$-\frac{\gamma_i}{N^2 \pi_i^{\,2}} + \frac{1}{N^2 n} \sum_{i=1}^{N} \frac{\gamma_i}{\pi_i} = 0 \qquad i = 1,\ldots,N \tag{IV.3a}$$

Let

$$\pi_i^* = \frac{n\sqrt{\gamma_i}}{\sum_{i=1}^{N} \sqrt{\gamma_i}} \quad .$$

Then π_i^* satisfies Equation IV.3a. That π_i^* minimizes the A-V(d_1) follows from the Kuhn-Tucker sufficiency theorem given in Hadley (19).

An alternative way to show the optimum property of π_i^* is to use the Hölder Inequality,

$$\sum_{i=1}^{N} |u_i v_i| \leq \left(\sum_{i=1}^{N} |u_i|^s\right)^{1/s} \left(\sum_{i=1}^{N} |v_i|^t\right)^{1/t}, \quad \frac{1}{s}+\frac{1}{t} = 1, s,t>0.$$

Letting

$$u_i = \sqrt{\frac{\gamma_i}{\pi_i}} \quad , \quad v_i = \sqrt{\pi_i}$$

we have

$$\Sigma|u_i v_i| = \sum_{i=1}^{N} \sqrt{\gamma_i} \leq \left(\sum_{i=1}^{N} \frac{\gamma_i}{\pi_i}\right)^{1/2} \left(\sum_{i=1}^{N} \pi_i\right)^{1/2} \quad .$$

This implies that

$$\frac{1}{n}\left(\sum_{i=1}^{N}\sqrt{\gamma_i}\right)^2 \leq \sum_{i=1}^{N}\frac{\gamma_i}{\pi_i} \quad .$$

The equality holds if $\pi_i = \pi_i^*$. Hence π_i^* minimizes the A-V(d_1).

Now, the lower bound result of Godambe and Joshi (17) assumes sampling designs utilizing the characteristic y only. However, with slight modification of the proof, their result can be extended to sampling of tuples (y,x) with $\bar{X}_N$ known. The modification of their theorem is given in Appendix 4. The A-V(d_1) satisfies this modified lower bound.

Theorem B.1 illustrates one interesting and somewhat intuitive result. Under the model, one should sample proportionately to the square root of the a priori variance. Since γ_i represents the measure of uncertainty of the survey designer about the closeness of y_i to x_i, the result in Theorem B.1 says to select units of which you are more uncertain with higher probability than those of which you have little doubt.

Corollary:

If $\gamma_i = \sigma^2$ for all i, then the estimator $\bar{y}_n + \bar{X}_N - \bar{x}_n$ together with SRS w/o replacement of size n are jointly optimum.

In conjunction with the results of Theorem B.1 and Model II we prove further properties of the estimator and design. Recall the definition of unbiased admissibility given in the beginning of the chapter. Let us assume that $x = (x_1,\ldots,x_N)$ is a fixed point in R^N whereas $y = (y_1,\ldots,y_N) \varepsilon R^N$ i.e., y is allowed to take any finite value in R^N. By using the same proof as Godambe and Joshi (17) with $s_k = \{y : \text{just } k\ y_i - x_i \neq 0,\ i = 1,\ldots,N\}$ it can be shown that

$$\sum_{i=1}^{n} \frac{y_i - x_i}{N\pi_i}$$

is unbiased admissible for $\bar{Y}_N - \bar{X}_N$.

Furthermore, suppose $\bar{X}_N$ is known prior to sampling. If

$$d_1 = \sum_{i=1}^{n} \frac{y_i - x_i}{N\pi_i} + \bar{X}_N$$

is not unbiased admissible for $\bar{Y}_N$, this implies there exists an estimator $f((y_i,x_i),\bar{X}_N)$ such that $V(f) \leq V(d_1)$ for all y with strict inequality for at least one y. Then $f - \bar{X}_N$ is such that

$$V(f - \bar{X}_N) = V(f) \leq V(d_1) = V\left(\sum_{i=1}^{n} \frac{y_i - x_i}{N\pi_i}\right) \quad .$$

We know that $d_1 - \bar{X}_N$ is unbiased admissible for $\bar{Y}_N - \bar{X}_N$. But $V(f-\bar{X}_N) \leq V(d_1 - \bar{X}_N)$ for all y with strict inequality for at least one y and hence we have a contradiction.

The unbiased admissibility of $d_1 - \bar{X}_N$ for $\bar{Y}_N - \bar{X}_N$ when $\bar{X}_N$ is known is outlined in Appendix 5. We have now proven the following proposition.

Proposition:

Given $\bar{X}_N$ with $(x_1, \ldots, x_N)$ a fixed point in R^N, draw a sample size n of tuples (y_i, x_i) $i = 1, \ldots, n$ with inclusion probabilities π_i,

$$\sum_{i=1}^{N} \pi_i = n \quad .$$

Then the estimator

$$d_1 = \sum_{i \varepsilon s} \frac{y_i}{N\pi_i} + \bar{X}_N - \sum_{i \varepsilon s} \frac{x_i}{N\pi_i}$$

is unbiased admissible for $\bar{Y}_N$.

Now, consider Model II with the restriction that we select a replacement sample of size n, the sample consisting of the characteristic y_i only. Given the estimator

$$d_2 = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i t_i}{np_i}$$

where t_i is the number of times the i^{th} unit appears in the sample. We wish to find the best p_i, $0 < p_i < 1$, $\sum^{N} p_i = 1$.

The

$$A\text{-}V(d_2) = V[\frac{1}{N}\sum_{i=1}^{N}\frac{x_i t_i}{np_i} - \bar{X}_N] + E\mathcal{V}[\frac{1}{N}\sum_{i=1}^{N}\frac{y_i t_i}{np_i} - \bar{Y}_N]$$

$$= \frac{1}{N^2 n}[\sum_{i=1}^{N} p_i(\frac{x_i}{p_i} - X_N)^2] + E[\frac{1}{N^2}\sum_{i=1}^{N}\frac{t_i^{\,2}\mathcal{V}_i}{(np_i)^2}$$

$$- \frac{2}{N^2}\sum_{i=1}^{N}\frac{\mathcal{V}_i t_i}{np_i} + \sum_{i=1}^{N}\frac{\mathcal{V}_i}{N^2}]$$

$$= \frac{1}{N^2 n}\sum_{i=1}^{N}\mathcal{V}_i - \frac{X_N^{\,2}}{N^2 n} + \frac{1}{N^2}\sum_{i=1}^{N}\frac{(\mathcal{V}_i + x_i^{\,2})}{np_i} \quad .$$

Using Hölder's inequality, the optimum p_i is proportional to $\sqrt{\mathcal{V}_i + x_i^{\,2}}$. At first glance the condition of with replacement sampling may diminish the appeal of this result. Consider the H.T. estimator for a fixed sample size w/o replacement design.

The $A-V\left(\sum_{i=1}^{n}\frac{y_i}{N\pi_i}\right)$

$$= \frac{1}{N^2}[\sum\frac{x_i^{\,2}}{\pi_i} + \sum_{i\neq j} x_i x_j \frac{\pi_{ij}}{\pi_i\pi_j} + X_N^{\,2}] + \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathcal{V}_i}{\pi_i}$$

$$- \sum_{i=1}^{N}\frac{\mathcal{V}_i}{N^2} \quad .$$

Hence for w/o replacement sample designs such that π_{ij} is approximately proportional to $\pi_i\pi_j$, $i\neq j$, then $\pi_i = np_i$ where p_i is proportional to $\sqrt{\mathcal{V}_i + x_i^{\,2}}$ v_i are the optimum inclusion probabilities.

The following two results are due to Godambe and Joshi (17). They assume knowledge of the x_i, $i=1,\ldots,N$ prior to sampling but do not assume knowledge of γ_i/γ_j, $i \neq j$. They also assume that the $(y_1,\ldots,y_N)$ are distributed independently with respect to some distribution. We then have

Theorem B.2 (Godambe and Joshi, (17)):

Given

i) Model II with $x_i > 0 \ \forall_i$

ii) w/o replacement sampling designs of fixed sample size n

iii) $\pi_i = \dfrac{nx_i}{\sum\limits^{N} x_i} < 1$ and

iv) we wish to consider unbiased estimators of $\bar{Y}_N$.

Then, the estimator

$$\frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

satisfies iv) and achieves the lower bound of Godambe and Joshi (17).

Proof:

The reader is referred to Godambe and Joshi (17) for a proof.

Further, if $\gamma_i = cx_i^2 \ \forall_i$ we have the corollary:

Corollary (Godambe and Joshi, (17)):

Under Model II with $\gamma_i = cx_i^{\,2}$, the estimator

$$\frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

together with the sampling design

$$\pi_i = \frac{nx_i}{\sum_{i=1}^{N} x_i} < 1$$

jointly satisfy the lower bound of Godambe and Joshi.

Proof:

The proof is given in Godambe and Joshi (17). The corollary is also a special case of Theorem B.1.

In summary, under Model II, the unequal probability difference estimator together with any design such that

$$\pi_i = \frac{n\sqrt{\gamma_i}}{\sum_{i=1}^{N} \sqrt{\gamma_i}} < 1$$

and $\bar{X}_N$ known are jointly optimum in the class of unbiased estimators and sampling designs with expected sample size equal to n.

C. Model III

Model III (Hajék, 20) is given by

$$\mathscr{E}(y_i|x_i) = x_i \quad , \qquad x_i > 0$$

$$\mathscr{C}ov(y_i,y_j|x_i,x_j) = cx_i^{\;2} \qquad i = j$$

$$= \sigma^2 x_i x_j R(|i-j|) \qquad i \neq j$$

where R is a convex function, c and σ^2 are unknown.

This model is due to Hajék (20) and is quite general in its covariance structure but is restrictive in that the anticipated means are positive and the anticipated variance is proportional to the square of the anticipated mean for all i.

Theorem C.1 (Hajék, 20):
Under Model III, in the class of sample designs of fixed size n and unbiased estimators in the class T_5, unequal probability systematic sampling with

$$\pi_i = \frac{nx_i}{\sum\limits_{i=1}^{N} x_i}$$

and the estimator given by

$$\frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

are jointly best for $\bar{Y}_N$ with respect to the A-V.

Proof:

The proof of this result is given in Hájek (20).

Although the class T_5 does not include all of the unbiased estimators it does include most of the estimators used in practice. For example, regression and ratio estimators belong to the class T_5.

D. Model IV

$$\mathscr{E}(y_i|x_i) = x_i$$

$$\mathscr{C}ov(y_i,y_j|x_i,x_j) = \gamma_i \qquad i = j$$

$$= \rho_{ij} \qquad i \neq j \qquad |\gamma| > 0 \quad .$$

Assume that we know x_i, γ_i, ρ_{ij} prior to sampling. It would be reasonable to find the estimator and sampling design of fixed size n such that the estimator is unbiased and its A-V is minimized over a class of estimators and sampling designs. We state the following proposition.

Proposition:

Under Model IV, the best unbiased estimator in the class T_5 and the best sampling design of size n may be found as the solution of the following programming problem:

Minimize

$$A\text{-}V(t_5) = \sum_s\Big(\sum_{i\varepsilon s} \beta_{is}x_i - \sum^N \frac{x_i}{N}\Big)^2 p(s) + \sum p(s)\Big\{\sum_{i\varepsilon s} \beta_{is}^2 \gamma_i$$

$$+ 2\sum_{ij,\varepsilon s} \beta_{is}\beta_{js}\rho_{ij} + \sum \frac{\gamma_i}{N^2} + \frac{1}{N^2}\sum_{i\neq j}^{N} \rho_{ij}$$

$$- \frac{2}{N}\Big(\sum_{i\varepsilon s} \beta_{is}\gamma_i + \sum_{\substack{i\varepsilon s\\ j\not\varepsilon s}} \beta_{is}\rho_{ij}\Big)\Big\} \quad .$$

Subject to

i) $\sum_{s\ni i} \beta_{is}p(s) = \frac{1}{N} \forall_i$

ii) $\sum_{s\varepsilon S} p(s) = 1$

iii) $p(s) \geq 0 \quad \forall s.$

While the proposition makes known the possibility of exact solutions in some cases, particularly for small populations, a programming solution is impractical for large populations due to the prohibitive cost involved.

E. Model V

Let

$$\mathcal{E}(y_i|x_i) = x_i + c$$

$$\mathcal{C}ov(y_i,y_j|x_i,x_j) = 0 \qquad i\neq j$$

$$= \gamma_i \qquad i = j$$

where c is unknown and $|\gamma_i| < \infty \; \forall_i$.

As in Model II, we shall assume that a sample of size n will consist of tuples (y_i, x_i). Furthermore, assume that $\bar{X}_N$ is known prior to sampling. First, we have

Theorem E.1:

Given Model V with $\gamma_i = \gamma_j$ for all $i \neq j$. Also given that $(y_1, \ldots, y_N)$ are independent with respect to some distribution. Then in the class of unbiased estimators and fixed sample size designs the difference estimator $\bar{y}_n + (\bar{X}_N - \bar{x}_n)$ together with SRS without replacement of size n satisfy the modified lower bound of Godambe and Joshi (17) (Appendix 5).

Proof:

The proof is completed by noting that the difference estimator is unbiased for $\bar{Y}_N$ under SRS w/o replacement of size n. Also, the A-V of the difference estimator satisfies the modified lower bound given in Appendix 4.

We now widen the class of possible estimators to include consistent estimators of $\bar{Y}_N$. Accordingly, the following results will be applicable only to the large sample situation.

In Appendix 8 we state without proof several theorems to be used in the remainder of the dissertation. For the present we define two types of order in probability.[1]

Def. The sequence of random variables $\langle X_n \rangle$ converges in probability to the random variable X and we write $\text{plim } X_n = X$ or $X_n \xrightarrow{P} X$ if for every $\varepsilon > 0$ and $\delta > 0$, there exists an N such that $\forall\, n > N$, $P[|X_n - X| \geq \varepsilon] < \delta$.

Def. X_n is of probability order $o_p(g_n)$ and we write $X_n = o_p(g_n)$ if

$$\text{plim } \frac{X_n}{g_n} = 0 \quad .$$

Def. X_n is of probability order $O_p(g_n)$ and we write $X_n = O_p(g_n)$ if for every $\varepsilon > 0$ there exists a positive real number M_ε such that

$$P\left[\left|\frac{X_n}{g_n}\right| \geq M_\varepsilon \right] \leq \varepsilon \quad \forall_n \quad .$$

Hence if $\varepsilon > 0$ and $X_n = O_p(\frac{1}{n^\varepsilon})$ then $X_n \xrightarrow{P} 0$.

Def. Given a sampling design p, an estimator a_n of $\bar{Y}_N$ is consistent for $\bar{Y}_N$ if $a_n \xrightarrow{P} \bar{Y}_N$.

In the following we give a way to establish consistency for estimators of the mean computed for a sample of size n selected without replacement with unequal probabilities. We

---

[1] Mann, H. B. and Wald, A. (35, p. 218).

shall concern ourselves solely with the estimator

$$\frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_i} \quad .$$

First, we shall summarize the results for replacement sampling of size n.

If a sample of size n is selected with replacement with selection probabilities p_i where

$$\sum_{i=1}^{N} p_i = 1 \quad ,$$

the usual estimator

$$e_1 = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{np_i}$$

has variance

$$V(e_1) = \frac{1}{n} \sum_{i=1}^{N} \frac{p_i}{N^2} (\frac{y_i}{p_i} - Y_N)^2 \quad .$$

Fixing (y_i, p_i) $i = 1, \ldots, N$ and N,

$$V(e_1) = \frac{C}{n}$$

where

$$C = \frac{1}{N^2} \sum_{i=1}^{N} p_i (\frac{y_i}{p_i} - Y_N)^2$$

is not dependent on n. Hence as n increases $V(e_1)$ decreases and by virtue of the Tchebychev inequality e_1 converges in probability to $\bar{Y}_N$.

In order to establish consistency for unequal probability non-replacement sampling designs of size n we need to consider several problems which do not arise in with replacement sampling.

The first problem is that of the population size N in relation to the sample size n. Since the sample consists of n distinct units, increasing n with N fixed implies that we would eventually examine all of the population. The survey design would then become a census. To avoid such a situation we can either

a) Let N and n increase in such a way that $\frac{n}{N}$ remains a fixed number, or

b) Let N and n increase in such a way that $\lim_{n,N\to\infty} \frac{n}{N} = f$, $0 < f < 1$, a fixed number.

With unequal probability w/o replacement sampling it is impossible to increase the sampling fraction and also maintain the relative magnitudes of the selection probabilities, p_i. This is because with $\pi_i = np_i$, increasing n necessitates decreasing p_i to maintain $\pi_i < 1$, as a valid probability.

If we permit the population size to increase we must specify the limiting behavior of the parameters of the population. We can restrict $\bar{Y}_N$ to remain fixed as N increases or we can merely stipulate that it be finite for all N.

The third problem is that of controlling the probabilities of selection p_i, $i = 1,\ldots,N$, $0 < p_i < 1$, as N increases. This problem is connected with the fourth problem which is to specify a design that yields n distinct units and an estimator with variance converging to zero in some sense as n gets large.

Suppose we are given $\pi_i = np_i$ where

$$\sum_{i=1}^{N} p_i = 1$$

and we are interested in obtaining n distinct units in a sample of size n drawn with unequal probability, p_i. To achieve this we group the units into n strata such that

$$\sum_{j=1}^{N_{(i)}} \pi_{(i)j} = 1$$

where $N_{(i)}$ is the stratum size and $\pi_{(i)j}$ is the inclusion probability of unit j in the i^{th} stratum. In our theoretical discussion we assume that such a grouping is possible with no rounding error. We note in passing that there are without replacement sampling designs of size n that maintain the inclusion probabilities π_i when the units cannot be formed into groups of unit probability, e.g.

Fuller (15). Having stratified the population into n strata, $N_{(i)}$ elements in stratum i,

$$\sum_{i=1}^{n} N_{(i)} = N ,$$

we obtain our sample of n distinct units by independently selecting one element per stratum with probability, $\pi_{(i)j}$.

The index (i) designates the stratum and j denotes the element within the i^{th} stratum. To estimate $\bar{Y}_N$, we use the unbiased estimator,

$$e_2 = \frac{1}{N} \sum_{i=1}^{n} \frac{y_{(i)j}}{\pi_{(i)j}} \tag{IV.3b}$$

The variance of this estimator is

$$V(e_2) = \frac{1}{N^2} \sum_{i=1}^{n} \sum_{j=1}^{N_{(i)}} \pi_{(i)j} \left(\frac{Y_{(i)j}}{\pi_{(i)j}} - Y_{(i)}\right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{N_{(i)}}{N}\right)^2 \frac{1}{N_{(i)}^{\,2}} \sum_{j=1}^{N_{(i)}} \pi_{(i)j} \left(\frac{Y_{(i)j}}{\pi_{(i)j}} - Y_{(i)}\right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{N_{(i)}}{N}\right)^2 C_{(i)}$$

where $C_{(i)}$ is the variance of

$$\frac{Y_{(i)j}}{N_{(i)}\pi_{(i)j}}$$

as an estimator of the stratum mean $\bar{Y}_{(i)}$, and accordingly is of order one. This designation of order follows as a consequence of the variance of the sample mean

$$\frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{np_i} ,$$

when sampling with unequal probabilities and with replacement, which is known to be of order $\frac{1}{n}$.

For the moment, we shall increase n and N in such a way that $\frac{n}{N} = f$, $0 < f < 1$, where f is a fixed number. We shall define an increasing sequence of populations, indexed by the sample size, where the increasing sequence of populations is constructed by increasing strata in proportion to the increase in n. For every unit increase in n, the population is increased by $\frac{1}{f}$ units, thereby maintaining the sampling fraction. (We assume $\frac{1}{f}$ is integral.) Hence, as n increases by 1 to $n+1$, $N = N_n$ increases to

$$N_{n+1} = \sum_{i=1}^{n+1} N_{(i)}$$

where

$$N_{(n+1)} = \frac{1}{f} \quad .$$

iii) Let $y_{(n+1)j}$ and $\pi_{(n+1)j}$ denote the population characteristic and inclusion probability of unit j in the $(n+1)^{st}$ stratum where

$$\sum_{j=1}^{N_{(n+1)}} \pi_{(n+1)j} = 1 \quad .$$

We shall assume that the population mean remains finite and

$$V\left(\frac{Y_{(n+1)j}}{N_{(n+1)}\pi_{(n+1)j}}\right) = \frac{1}{N^2_{(n+1)}} \sum_{j=1}^{N_{(n+1)}} \pi_{(n+1)j}\left(\frac{Y_{(n+1)j}}{\pi_{(n+1)j}} - Y_{(n+1)}\right)^2$$

$$= C_{(n+1)} < C \quad \forall n \quad .$$

Hence the variance of

$$\frac{Y_{(n+1)j}}{N_{(n+1)}\pi_{(n+1)j}}$$

when one unit is selected with probability $\pi_{(n+1)j}$ in stratum (n+1) is bounded by some number C for all n. Sample selection in the additional strata is performed independently from stratum to stratum, one unit being selected per stratum. As n, N increase the inclusion probabilities of the original N units do not change. In general, the estimator in IV.3b is

$$e_2 = \frac{1}{N_{n'}} \sum_{i=1}^{n'} \frac{Y_{(i)j}}{\pi_{(i)j}}$$

with variance

$$V(e_2) = \sum_{i=1}^{n'} \left(\frac{N_{(i)}}{N_{n'}}\right)^2 C_{(i)}$$

and where n' denotes the sample size.

Consequently, since $n' = fN_{n'}$ $\forall$ $n' > n$,

$$V(e_2) \leq (\text{Constant})\frac{f}{n'} \sum_{i=1}^{n'} \frac{N_{(i)}^2}{N_{n'}} \leq \frac{\text{Constant}}{N_{n'}} \quad .$$

It follows that

$$\lim_{n'\to\infty} V(e_2) = 0 \quad .$$

Hence e_2 is a consistent estimator of $\bar{Y}_N$ by virtue of the Tchebychev inequality.

The one unit per stratum survey design does not allow an unbiased estimator of variance. To achieve this we group the N elements into n/2 strata in such a way that the sum of the given inclusion probabilities of the units in any stratum sum to 2. Here again, we shall neglect any rounding error due to grouping. To select n distinct units we select two elements per stratum with unequal probability and without replacement. Selection of pairs of elements is independent over strata. Denote by $\pi_{(i)j}$ the inclusion probability of element j in stratum i,

$$\sum_{j=1}^{N_{(i)}} \pi_{(i)j} = 2 \quad ,$$

and let $\pi_{(i)jk}$ denote the inclusion probability of units j and k in stratum i. There are a number of sampling designs that may be utilized in selecting 2 units per stratum. For example, Durbin (9) has a design in which

$$\pi_{(i)jk} = \frac{\pi_{(i)j}\pi_{(i)k}}{2(1+A_{(i)})}\left[\frac{1}{1-\pi_{(i)j}} + \frac{1}{1-\pi_{(i)k}}\right]$$

where

$$A_{(i)} = \frac{1}{2}\sum_{j=1}^{N_{(i)}} \frac{\pi_{(i)j}}{1-\pi_{(i)j}} \quad .$$

Correspondingly, the estimator

$$e_3 = \frac{1}{N}\sum_{i=1}^{n/2} \left(\frac{y_{(i)j}}{\pi_{(i)j}} + \frac{y_{(i)k}}{\pi_{(i)k}}\right) \tag{IV.3c}$$

is unbiased for the population mean and has a variance

$$V(e_3) = \sum_{i=1}^{n/2} \left(\frac{N_{(i)}}{N}\right)^2 D_{(i)} \tag{IV.3d}$$

where

$$D_{(i)} = \frac{1}{N_{(i)}^2}\left[\sum_{j\neq k}^{N_{(i)}} (\pi_{(i)j}\pi_{(i)k} - \pi_{(i)jk})\left\{\frac{y_{(i)j}}{\pi_{(i)j}} - \frac{y_{(i)k}}{\pi_{(i)k}}\right\}^2\right].$$

In our discussion we shall assume the following:

i) The population size N and sample size n are assumed to increase in such a way that

$$\lim_{n\to\infty} f_n = f < 1$$

where $f_n = n/N$ and n increases by increments of 2 units.

ii) As in the previous development, let n,N be given. Then

$$N_n = \sum_{i=1}^{n/2} N_{(i)}$$

and a sequence of populations is constructed by increasing strata so that for each increase of 2 units in n, $n+2 = f_{n+2}N_{n+2}$ where

$$N_{n+2} = \sum_{i=1}^{n+2/2} N_{(i)}$$

and $N_{((n+2)/2)}$ is the number of units in the appended stratum.

iii) In addition, the stratum means are bounded and the probabilities of inclusion for each stratum sum to 2.

iv) The term $D_{(i)}$ given in Equation IV.3d is bounded for all i,

$$D_{(i)} < D < \infty \quad \forall_i \quad .$$

Consider estimator e_3 given in Equation IV.3c. For a given n, the

$$V(e_3) = \sum_{i=1}^{n/2} \left(\frac{N_{(i)}}{N}\right)^2 D_{(i)} \quad .$$

Since $D_{(i)}$ is bounded for all i,

$$V(e_3) \leq D \sum_{i=1}^{n/2} \left(\frac{N_{(i)}}{N}\right)^2 = D \sum_{i=1}^{n/2} \left(\frac{f_n N_{(i)}}{n}\right)^2 \quad .$$

Let

$$\max_{1 \leq i \leq n} (f_n N_{(i)})$$

be bounded by some constant C'. That such a C' exists follows from the fact that

$$\lim_{n\to\infty} f_n = f < 1 \quad .$$

For sufficiently large n each increase of 2 units in n will result in appending a new stratum of approximately $\frac{2}{f}$ units. Hence

$$V(e_3) \leq \frac{DC'}{2n} \qquad \text{(IV.3e)}$$

and e_3 is consistent for $\bar{Y}_N$.

In summary, the H.T. estimator is a consistent estimator for the population mean in the case of sampling w/o replacement with n distinct units from an increasing sequence of populations constructed to contain n/2 strata. Furthermore, under this structure, the sequence of populations is such that the original inclusion probabilities remain unchanged as the size of the population increases.

Consider Model V and assume that $\bar{X}_N$ is known. Further, assume that we wish to select with unequal probability and without replacement a sample of n tuples (y,x). Let π_i denote the inclusion probability of unit i. The variance of the Horwitz-Thompson estimator was given in Equation IV.3e and was shown to be of $O(\frac{1}{n})$. It follows that

$$\frac{1}{N}\sum_{i=1}^{n}\frac{z_i}{\pi_i} - \bar{Z}_N = o_p\left(\frac{1}{\sqrt{n}}\right) \qquad \text{(IV.4)}$$

where z_i is any characteristic related to unit i.

One way to proceed given Model V is to estimate the unknown translation parameter c and use the estimator $\bar{X}_N + \hat{c}$ where $\hat{c}$ is an estimator of c based on the sample. Assuming the infinite model given by Model V and estimating c by least squares, we have

$$\hat{c} = \sum^{n} \frac{y_i - x_i}{\mathcal{V}_i} \Big/ \sum^{n} \frac{1}{\mathcal{V}_i} \quad . \qquad \text{(IV.5)}$$

The estimator $\bar{X}_N + \hat{c}$ is unbiased under the model, since

$$\mathcal{E}(\bar{X}_N + \hat{c}|n) = \bar{X}_N + c \sum^{n} \frac{1}{\mathcal{V}_i} \Big/ \sum^{n} \frac{1}{\mathcal{V}_i} = \bar{X}_N + c \quad .$$

Furthermore, under the model, $\hat{c}$ is the best linear unbiased estimator of c; that is, of all estimators of the form

$$\sum_{i=1}^{n} \beta_{is} y_i + f(s)$$

where f is a function of s but not of the $y_i \varepsilon s$, $\mathcal{V}(\hat{c}|n)$ is smallest. This statement is formalized as Theorem E.2 and a proof is given in Appendix 7.

Let us generalize $\hat{c}$ somewhat and denote by $\hat{c}'$ the estimator

$$\hat{c}' = \sum^{n} \frac{y_i - x_i}{\pi_i} \Big/ \sum^{n} \frac{1}{\pi_i} \quad .$$

Notation:

In the remainder of the dissertation we will need to calculate expectations of ratios of random variables. When calculating these expectations we will expand the ratio in a Taylor's series and omit the expectation of terms of specified order in probability. The expectation of the resulting terms will be denoted by $\bar{E}$ instead of E. The designation of the resulting expression will also be denoted with a raised bar ($^{-}$).

Consider the estimator

$$\bar{X}_N + \hat{c}' = \bar{X}_N + \frac{1}{N} \sum^{n} \frac{y_i - x_i}{\pi_i} \Big/ \frac{1}{N} \sum^{n} \frac{1}{\pi_i} \quad .$$

Utilizing the order in probability of the Horwitz-Thompson estimator and Theorem I in Appendix 8, it is easy to see that $\bar{X}_N + \hat{c} \xrightarrow{P} \bar{Y}_N$.

Under Model V,

$$y_i = x_i + c + e_i$$

where

$$\mathcal{E}(e_i) = 0$$

$$\begin{aligned} \mathcal{C}ov(e_i, e_j) &= \gamma_i \qquad i = j \\ &= 0 \qquad i \neq j \end{aligned}$$

and where c is unknown. Assume that all inclusion probabilities π_i are bounded, i.e., for some $\varepsilon > 0$,

$$\varepsilon < \pi_i < 1 - \varepsilon \quad \forall_i \quad .$$

Also, assume that all $\mathscr{V}_i$ are bounded. We now find the A-V$(\bar{X}_N + \hat{c}')$ omitting terms of order in probability larger than $\frac{1}{n}$. The estimator $\bar{X}_N + \hat{c}'$ is unbiased under the model since

$$\mathscr{E}(\bar{X}_N + \hat{c}' - \bar{Y}_N | s) = \bar{X}_N + c - \bar{X}_N - c = 0 \quad .$$

Thus

$$AV(\bar{X}_N + \hat{c}') = E\mathscr{V}(\bar{X}_N + \sum^n \frac{y_i - x_i}{N\pi_i} / \sum^n \frac{1}{N\pi_i} - \bar{Y}_N | s) \quad . \qquad (IV.6)$$

Substituting $y_i = x_i + c + e_i$ into Equation IV.6 we have

$$A\text{-}V(\bar{X}_N + \hat{c}') = E\mathscr{V}[\sum^n \frac{e_i}{N\pi_i} / \sum^n \frac{1}{N\pi_i} - \bar{e}_N | s]$$

$$= E[D^2 \sum^n \frac{\mathscr{V}_i}{N^2 \pi_i^{\,2}} - 2D \sum^n \frac{\mathscr{V}_i}{N^2 \pi_i} + \sum_{i=1}^N \frac{\mathscr{V}_i}{N^2}] \qquad (IV.7)$$

where

$$D = \frac{1}{\sum^n \frac{1}{N\pi_i}} \quad .$$

Let us examine the terms on the right hand side of Equation IV.7. Expanding D^2 in a Taylor's series we have

$$D^2 \sum_{i=1}^{n} \frac{\mathcal{V}_i}{N^2\pi_i^{\,2}} = \sum_{i=1}^{n} \frac{\mathcal{V}_i}{N^2\pi_i^{\,2}}(1 + 2(1 - \sum^{n} \frac{1}{N\pi_i}) + O_p(\frac{1}{n}))$$

$$= \sum^{n} \frac{\mathcal{V}_i}{N^2\pi_i^{\,2}} + 2(1 - \sum^{n} \frac{1}{N\pi_i})\sum^{n} \frac{\mathcal{V}_i}{N^2\pi_i^2} + \frac{1}{N}\, Op(\frac{1}{n}) \sum_{i=1}^{n} \frac{\mathcal{V}_i}{N\pi_i^{\,2}} \quad .$$

In the sequence of populations given earlier we assumed that

$$\lim_{n\to\infty} \frac{n}{N} = f < 1 \quad .$$

Hence any term that is of order $\frac{1}{n}$ is also of order $\frac{1}{N}$. For example,

$$\frac{1}{nN} = O(\frac{1}{n^2}) = O(\frac{1}{N^2}) \quad .$$

Now,

$$D^2 \sum^{n} \frac{\mathcal{V}_i}{N^2\pi_i^{\,2}} = \frac{1}{N}[\sum^{n} \frac{\mathcal{V}_i}{N\pi_i^{\,2}} + 2(1 - \sum^{n} \frac{1}{N\pi_i})(\sum \frac{\mathcal{V}_i}{N\pi_i^{\,2}} - \frac{\sum^{N} \mathcal{V}_i}{N\pi_i})$$

$$+ 2(1 - \sum^{n} \frac{1}{N\pi_i}) \sum_{i=1}^{N} \frac{\mathcal{V}_i}{\pi_i N} + O_p(\frac{1}{n^2}) \sum^{n} \frac{\mathcal{V}_i}{N\pi_i^{\,2}}] \quad .$$

By assumption $\mathcal{V}_i/\pi_i$ is bounded for every i. From Equation IV.4, we have

$$\frac{1}{N} \sum^{n}(\mathcal{V}_i/\pi_i)\frac{1}{\pi_i} - \frac{1}{N} \sum_{i=1}^{N} \mathcal{V}_i/\pi_i = O_p(\frac{1}{\sqrt{n}}) \quad .$$

It follows that omitting terms of order in probability larger than 1/n,

$$D^2 \sum^n \frac{\gamma_i}{N^2\pi_i^2} = \frac{1}{N}\left[\sum^n \frac{\gamma_i}{N\,\pi_i^2}\right] \tag{IV.8}$$

The second term on the right hand side of Equation IV.7 is

$$-\,2D \sum_{i=1}^n \frac{\gamma_i}{N^2\pi_i} \quad .$$

Expanding D in a Taylor's series we have

$$-2D \sum^n \frac{\gamma_i}{N^2\pi_i} = -\frac{2}{N} \sum^n \frac{\gamma_i}{N\pi_i}\left[1 + \left(1 - \sum^n \frac{1}{N\pi_i}\right) + O_p\left(\frac{1}{n}\right)\right] \quad .$$

Again, omitting terms of order in probability larger than 1/n, we have

$$-2D \sum^n \frac{\gamma_i}{N^2\pi_i} = -\frac{2}{N} \sum_{i=1}^n \frac{\gamma_i}{N\pi_i} \tag{IV.9}$$

Substituting Equations IV.8 and IV.9 into IV.7 we have

$$\overline{A\text{-}V}(\bar{X}_N + \hat{c}') = \bar{E}\left[\sum^n \frac{\gamma_i}{N^2\pi_i^2} - \frac{2}{N}\sum^n \frac{\gamma_i}{N\pi_i} + \frac{1}{N^2}\sum^N \gamma_i\right]$$

$$= \frac{1}{N^2}\sum_{i=1}^N \frac{\gamma_i}{\pi_i} - \frac{1}{N^2}\sum_{i=1}^N \gamma_i \quad . \tag{IV.10}$$

The expression given in IV.10 is exactly that given in section B. Hence utilizing the same minimizing technique used in section B, we have that

$$\pi_i = \frac{n\sqrt{\gamma_i}}{\sum^N \sqrt{\gamma_i}}$$

minimizes the $\overline{A\text{-}V}(\bar{X}_N + \hat{c}')$.

We now calculate the large sample variance of $\bar{X}_N + \hat{c}'$.

Define

$$C = \bar{Y}_N - \bar{X}_N$$

and let

$$y_i = C + x_i + u_i \quad .$$

It follows that

$$\sum^N u_i = 0 \quad .$$

We have

$$\bar{X}_N + \hat{c}' - \bar{Y}_N = \bar{X}_N + C + \sum^n \frac{u_i}{N\pi_i} / \sum^n \frac{1}{N\pi_i} - \bar{Y}_N$$

$$= \sum^n \frac{u_i}{N\pi_i} / \sum^n \frac{1}{N\pi_i} \quad .$$

Using a Taylor's series expansion,

$$\bar{X}_N + \hat{c}' - \bar{Y}_N = \sum^n \frac{u_i}{N\pi_i}(1 + (1 - \sum^n \frac{1}{N\pi_i}) + O_p(\tfrac{1}{n})) \quad . \tag{IV.11}$$

Squaring both sides of Equation IV.11 and omitting terms of order in probability larger than 1/n, we have

$$(\bar{X}_N + \hat{c}' - \bar{Y}_N)^2 = (\sum^n \frac{u_i}{N\pi_i})^2 \quad .$$

Taking expectation of both sides

$$\overline{\text{M.S.E.}}(\bar{X}_N + \hat{c}') = \bar{E}((\sum^n \frac{u_i}{N\pi_i})^2)$$

$$= \frac{1}{N^2} \sum_{i \neq j} (\pi_i \pi_j - \pi_{ij})(\frac{u_i}{\pi_i} - \frac{u_j}{\pi_j})^2$$

where $u_i = y_i - \bar{Y}_N - (x_i - \bar{X}_N)$ and π_{ij} denotes the inclusion probability of units i and j.

Let us summarize the properties of $\bar{X}_N + \hat{c}'$ given Model V. The estimator $\bar{X}_N + \hat{c}'$

i) is a consistent estimator of the population mean

ii) has $\overline{\text{A-V}}$ minimized for π_i proportional to $\sqrt{\gamma_i}$ and

iii) has

$$\overline{\text{M.S.E.}} = \frac{1}{N^2} \sum_{i \neq j}^{N} (\pi_i \pi_j - \pi_{ij})(\frac{u_i}{\pi_i} - \frac{u_j}{\pi_j})^2$$

where $u_i = y_i - x_i - C$.

F. Model VI

Let

$$\mathcal{E}(y_i | x_i) = \beta x_i$$

$$\mathcal{C}ov(y_i, y_j | x_i, x_j) = \mathcal{V}_i \qquad i = j$$

$$= 0 \qquad i \neq j$$

where β is unknown. Model VI is called the ratio model. As in Models II and V, we shall assume sampling of n tuples (y,x) with $\bar{X}_N \neq 0$, known. Sampling is assumed to be with unequal probability and without replacement of fixed size n.

Consider the estimator

$$\hat{R}\bar{X}_N = [\sum^n \frac{y_i}{N\pi_i} / \sum^n \frac{x_i}{N\pi_i}]\bar{X}_N \quad , \qquad \text{(IV.12)}$$

where π_i denotes the inclusion probability of unit i. We shall assume that

$$\sum^n \frac{x_i}{N\pi_i} \neq 0$$

for all samples s so that $\hat{R}\bar{X}_N$ is well defined.

In the discussion of Model V we alluded to the order in probability of the Horwitz-Thompson estimator. We stated that

$$\sum_{i=1}^n \frac{z_i}{N\pi_i} - \bar{Z}_N = O_p(\frac{1}{\sqrt{n}}) \quad .$$

In this section we use the same arguments as in section

E in Chapter IV to establish the $\overline{\text{A-V}}(\hat{R}\bar{X}_N)$.

Consider Model VI,

$$y_i = \beta x_i + e_i$$

where

$$\mathcal{E}(e_i) = 0$$

$$\mathcal{C}ov(e_i, e_j) = \mathcal{V}_i \qquad i = j$$

$$= 0 \qquad i \neq j$$

and where β is unknown. Assume that all inclusion probabilities π_i are bounded, i.e., for some $\varepsilon > 0$,

$$\varepsilon < \pi_i < 1 - \varepsilon \quad \forall_i \quad .$$

Also, assume that all $\mathcal{V}_i$ are bounded. We now find the $\text{A-V}(\hat{R}\bar{X}_N)$ omitting terms of order in probability larger than $1/n$. Since

$$\mathcal{E}(\hat{R}\bar{X}_N - \bar{Y}_N|s) = \beta\bar{X}_N + 0 - \beta\bar{X}_N = 0 \quad ,$$

$\hat{R}\bar{X}_N$ is unbiased under the model. The

$$\text{A-V}(\hat{R}\bar{X}_N) = E\mathcal{V}[\sum^n \frac{e_i}{N\pi_i} / \sum^n \frac{x_i}{N\pi_i} - \bar{e}_N|s]$$

$$= E[D^2 \sum^n \frac{\mathcal{V}_i}{(N\pi_i)^2} - 2D \sum^n \frac{\mathcal{V}_i}{N^2\pi_i} + \sum^N \frac{\mathcal{V}_i}{N^2}]$$

where

$$D = (\sum^n \frac{x_i}{N\pi_i})^{-1} \quad .$$

But this is the same form as Equation IV.7.

Hence the

$$\overline{A\text{-}V}(\hat{R}\bar{X}_N) = \overline{A\text{-}V}(\bar{X}_N + \hat{c}') = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\gamma_i}{\pi_i} - \sum^{N}\frac{\gamma_i}{N^2}$$

and the 'best' π_i is proportional to $\sqrt{\gamma_i}$.

We now find the M.S.E. $(\hat{R}\bar{X}_N)$ omitting expectation of terms of order in probability larger than $1/n$. Define

$$R_N = \frac{\bar{Y}_N}{\bar{X}_N}$$

and let

$$y_i = R_N x_i + s_i \quad . \tag{IV.13}$$

It follows that

$$\sum_{i=1}^{N} s_i = 0 \quad .$$

From Equation IV.13 we have

$$\hat{R}\bar{X}_N - \bar{Y}_N = \sum^{n}\frac{s_i}{N\pi_i}/\sum^{n}\frac{x_i}{N\pi_i} \quad .$$

Since the H.T. estimator

$$\sum^{n}\frac{z_i}{N\pi_i}$$

is consistent for $\bar{Z}_N$, using Theorem II in Appendix 8, $\hat{R}\bar{X}_N$ is a consistent estimator of $\bar{Y}_N$.

Expanding the left hand side in a Taylor's series and squaring we have

$$(\hat{R}\bar{X}_N - \bar{Y}_N)^2 = (\sum^n \frac{s_i}{N\pi_i})^2 \quad (1 + o_p(\frac{1}{\sqrt{n}})) \quad .$$

Omitting expectation of terms of order in probability larger than 1/n we have

$$\overline{\text{M.S.E.}}(\hat{R}\bar{X}_N) = \bar{E}[(\sum^n \frac{s_i}{N\pi_i})^2]$$

$$= \frac{1}{N^2} \sum_{i \neq j}^{N} (\pi_i \pi_j - \pi_{ij}) (\frac{s_i}{\pi_i} - \frac{s_j}{\pi_j})^2$$

where $s_i = y_i - R_N x_i$.

Summarizing the above results we have

Theorem F.1:

Given that we select a without replacement sample of fixed size n with inclusion probabilities π_i. The ratio estimator, $\hat{R}\bar{X}_N$ has the following properties,

i) it is a consistent estimator of $\bar{Y}_N$,

ii) its $\overline{\text{A-V}}$ is minimized for π_i proportional to $\sqrt{\gamma_i}$

iii) the

$$\overline{\text{M.S.E.}}(\hat{R}\bar{X}_N) = \frac{1}{N^2} \sum_{i \neq j} (\pi_i \pi_j - \pi_{ij}) (\frac{s_i}{\pi_i} - \frac{s_j}{\pi_j})^2$$

where $s_i = y_i - R_N x_i$.

1. Mixture type estimator 1

Def. Given any two estimators e_1 and e_2 of $\bar{Y}_N$ and a sampling design p, an estimator of the form

$$\hat{\theta} = \alpha(s)e_1 + \{1 - \alpha(s)\}e_2$$

where $\alpha(s)$ is a function of the sample s (hence possibly of $y_i \varepsilon s$) and $0 < \alpha(s) < 1$ will be called a mixture estimator of $\bar{Y}_N$.

Under Model VI and in the class of estimators T_5,

$$\hat{B} = \sum^n \frac{y_i x_i}{\mathcal{V}_i} / \sum^n \frac{x_i^2}{\mathcal{V}_i}$$

is the least squares estimator of β, i.e., given any sample s, $\mathcal{V}(\hat{B}|s)$ is smallest of all estimators $t_5 \varepsilon T_5$ where $\mathcal{E}(t_5|s) = \beta$.

However, the estimator $\hat{B}\bar{X}_N$ would not be considered by most survey designers. This is because $\hat{B}\bar{X}_N$ is not a consistent (with respect to the sampling design) estimator of $\bar{Y}_N$. In the following, we consider a mixture type estimator utilizing $\hat{B}\bar{X}_N$ and $\hat{R}\bar{X}_N$ that tends to the estimator $\hat{R}\bar{X}_N$ ($\alpha(s)$ tends to zero) when Model VI does not hold and tends to $\hat{B}\bar{X}_N$ ($\alpha(s)$ tends to one) when Model VI does hold.

Consider the following special case of Model VI given by

$$\mathcal{E}(y_i|s_i) = \beta x_i$$

$$\mathcal{C}ov(y_i, y_j | x_i, x_j) = \sigma^2 \qquad i = j$$

$$= 0 \qquad i \neq j$$

where β and σ^2 are unknown. We select a sample of size n of tuples (y,x) with equal probability and without replacement. Since we shall be interested in the large sample properties of estimators, we examine such properties with respect to a sequence of populations.

Accordingly, let

i) $pop_1, pop_2, \ldots, pop_n, \ldots$ denote a sequence of populations, symbolized by $\langle pop_n \rangle_{n=1}^{\infty}$,

ii) $f_1^{-1}1, f_2^{-1}2, \ldots, f_n^{-1}n, \ldots$ denote the corresponding population sizes,

iii) $\bar{Y}_{N_1}, \bar{Y}_{N_2}, \ldots, \bar{Y}_{N_n}, \ldots$ denote the corresponding means of the populations,

iv) $S^2(y)_{N_1}, S^2(y)_{N_2}, \ldots, S^2(y)_{N_n}, \ldots$ denote the corresponding population "S^2", and

v) $B_{N_1}, B_{N_2}, \ldots, B_{N_n}, \ldots$ denote the corresponding population regression coefficients, e.g.

$$B_{N_1} = \sum^{N_1} \frac{x_i y_i}{\sum^{N_1} x_i^2} ,$$

where all the above sequences are indexed by the sample size n and

vi) $\lim f_n = f$, $0 < f < 1$

$\lim \bar{Y}_{N_n} = \bar{Y}$, $|\bar{Y}| < \infty$

$\lim S^2(y)_{N_n} = S^2(y)$, $0 < S^2(y) < \infty$

$\lim B_{N_n} = B$, $|B| < \infty$

where all limits are with respect to n.

Let similar notation hold for $\bar{X}_{N_n}$, $S^2(x)_{N_n}$ and assume that $\bar{X}_{N_n}$ is known for all n.

For a particular population, say pop_n, we define the population characteristics by

$$R_{N_n} = \frac{\bar{Y}_{N_n}}{\bar{X}_{N_n}}$$

$$B_{N_n} = \frac{\sum^{N_n} y_i x_i}{\sum^{N_n} x_i^2}$$

and define s_i and u_i by

$$y_i = R_{N_n} x_i + s_i$$

and

$$y_i = B_{N_n} x_i + u_i \quad . \qquad \text{(IV.14)}$$

From the above definitions it follows that

$$\sum_{i=1}^{N_n} s_i = 0 = \sum^{N_n} u_i x_i \quad .$$

Let $\bar{y}_n$ be the sample mean. Then, given $\varepsilon > 0$, we have

$$P[\,|\bar{y}_n - \bar{Y}_{N_n}| \geq \varepsilon\,] \leq \frac{(1 - f_n)S^2(y)_{N_n}}{n\varepsilon^2}$$

by Tchebychev's inequality. With the specifications of vi) above, we have

$$\bar{y}_n - \bar{Y}_{N_n} = O_p(\frac{1}{\sqrt{n}})$$

and hence the sample mean is a consistent estimator of the population mean. By virtue of the sequence of populations, and the triangular inequality

$$P[\,|\bar{y}_n - \bar{Y}| \geq \varepsilon\,] \leq P[\,|\bar{y}_n - \bar{Y}_{N_n}| \geq \varepsilon/2\,] + P[\,|\bar{Y}_{N_n} - \bar{Y}| \geq \varepsilon/2\,] \,.$$

From vi) above,

$$\lim_{n\to\infty} \bar{Y}_{N_n} = \bar{Y} \,.$$

Hence given $\varepsilon > 0$, there exists N_ε such that for all $n \geq N_\varepsilon$ $|\bar{Y}_{N_n} - \bar{Y}| < \varepsilon$ with probability one. It follows that since

$$\bar{y}_n - \bar{Y}_{N_n} = O_p(\frac{1}{\sqrt{n}})$$

for n large,

$$P[\,|\bar{y}_n - \bar{Y}| \geq \varepsilon\,] \leq P\,[\,|\bar{y}_n - \bar{Y}_{N_n}| \geq \varepsilon/2\,] \quad .$$

The right hand side of the above inequality can be made arbitrarily small. Hence $\bar{y}_n$ converges in probability to

$\bar{Y}$. We maintain that consistency of estimators under a without replacement design of size n must necessarily be with respect to a sequence of populations. This is because of the requirement that $n \leq N$. Of course, under simple random sampling w/o replacement, it is possible to obtain

$$P[|\bar{y}_n - \bar{Y}_N| \geq \varepsilon] \leq (1 - \tfrac{n}{N}) \frac{S^2(y)_N}{n\varepsilon^2}$$

for a given $\varepsilon > 0$ without considering a sequence of populations. However, for very small $\varepsilon > 0$, this would require a sample size equal to the population size.

Sufficient conditions on $\langle pop_n \rangle_{n=1}^{\infty}$ (in addition to that of vi) above) to insure that $\bar{y}_n$ converges in prcbability to $\bar{Y}$ are

i) that pop_n have the same moments for all integers n and that $\lim f_n = f$ or

ii) that the pop_n is itself a random sample of size N_n from a universe with the properties given as the limits in vi).

Consider the two estimators

$$\hat{R}\bar{X}_{N_n} = \frac{\sum^{n} y_i}{\sum^{n} x_i} \bar{X}_{N_n}$$

and

$$\hat{B}\bar{X}_{N_n} = \frac{\sum^{n} x_i y_i}{\sum^{n} x_i^2} \bar{X}_{N_n}$$

where $\bar{X}_{N_n}$ is known prior to sampling. Assume that

$$\sum_{i=1}^{n} x_i \neq 0$$

for all $\binom{N}{n}$ samples s so that $\hat{R}\bar{X}_{N_n}$ is well defined. Henceforth we shall omit the subscript n in N_n to simplify notation. Before proceeding further we have

Theorem F.2:

Given that we sample n tuples (y_i, x_i) with SRS without replacement with $\bar{X}_N$ known. Then, omitting expectation of terms of order in probability larger than 1/n, we have

$$\text{i)}\quad \bar{V}(\hat{B}) = \frac{1}{\left(\Sigma \frac{N x_i^2}{N}\right)^2} \frac{1}{n}\left(1 - \frac{n}{N}\right) \sum^{N} \frac{(y_i - B_N x_i)^2 x_i^2}{N-1},$$

$$\text{ii)}\quad \bar{V}(\hat{R}) = \frac{1}{\bar{X}_N^2} \frac{1}{n}\left(1 - \frac{n}{N}\right) \sum^{N} \frac{(y_i - R_N x_i)^2}{N-1}$$

$$\text{iii)}\quad \overline{\text{Cov}}(\hat{B}, \hat{R}) = \frac{1}{\bar{X}_N \frac{\sum^{N} x_i^2}{N}} \frac{1}{n}\left(1 - \frac{n}{N}\right) \sum^{N} \frac{(y_i - B_N x_i)(y_i - R_N x_i) x_i}{N-1}$$

$$\text{iv)}\quad \bar{V}(\hat{B} - \hat{R}) = \bar{V}(\hat{B}) + \bar{V}(\hat{R}) - 2\overline{\text{Cov}}(\hat{B}, \hat{R}) \quad .$$

Furthermore, omitting expectation of terms of order in probability larger than $1/\sqrt{n}$ we have

$$\text{v)}\quad \bar{E}(\hat{B}) = B_N \quad \text{and}$$

vi) $\bar{E}(\hat{R}) = R_N$.

Proof:

The proof of Theorem F.2 follows by expanding $\hat{B}-B_N$ and $\hat{R}-R_N$ in a Taylor's series and noting that the resulting functions consist of sample means. The order in probability of the terms in the expansion may be established by Theorem II in Appendix 8. For example, from Equation IV.14 we have

$$\hat{B} - B_N = \frac{\sum^n (x_i u_i)}{\sum^n x_i^2} .$$

Squaring $\hat{B}-B_N$ and taking expectations we have

$$E[(\hat{B} - B_N)^2] = E[(\sum^n \frac{x_i u_i}{n})^2 / (\sum^n \frac{x_i^2}{n})^2] .$$

Expanding

$$1/\sum^n \frac{x_i^2}{n}$$

about the point

$$1/\sum^N \frac{x_i^2}{N} ,$$

yields

$$\bar{E}[(\hat{B}-B)^2] = \frac{1}{(\sum^N \frac{x_i^2}{N})^2} \bar{E}[(\sum^n \frac{x_i u_i}{n})^2 (1 - 2(\frac{\sum^n \frac{x_i^2}{n} - \sum^N \frac{x_i^2}{N}}{\sum^N \frac{x_i^2}{N}}) + o_p(\frac{1}{n}))]$$

$$= \frac{1}{\left(\sum^{N} \frac{x_i^2}{N}\right)^2} \frac{1}{n}\left(1 - \frac{n}{N}\right) \sum_{i=1}^{N} \frac{x_i^2 (y_i - B_N x_i)^2}{N-1} .$$

In the following we shall consider limiting distributions of sample statistics utilizing the results of Madow (33). To do this, it will be sufficient to place the following assumption on the sequence of populations. This assumption is called Condition W by Madow (33).

Assumption (Condition W):

Let

i) $$S^r(y)_N = \sum^{N} \frac{(y_i - \bar{Y}_N)^r}{N-1}$$

$$S^r(x)_N = \sum^{N} \frac{(x_i - \bar{X}_N)^r}{N-1}$$

$$S^r(x,y)_N = \sum^{N} \frac{\left(x_i y_i - \sum^{N} \frac{x_i y_i}{N}\right)^r}{N-1}$$

$$\rho_N = S^2(x,y)_N / [S^2(x)_N S^2(y)_N]^{1/2} .$$

Then, assume that the sequence of populations, $\langle pop_n \rangle$, is such that

ii) $$S^r(y)_N = [S^2(y)_N]^{r/2} \lambda_r(N)$$

$$S^r(x)_N = [S^2(x)_N]^{r/2} \lambda_r(N)$$

$$S^r(x,y)_N = [S^2(x,y)_N]^{r/2} \lambda_r(N)$$

for sufficiently large n and N, where a finite value λ exists such that for all r,

$$\sup_N |\lambda_r(N)| < \lambda \quad \text{and}$$

iii) $\lim\limits_{n\to\infty} \rho_N = \rho, \ |\rho| < 1$.

The above assumption in conjunction with Theorem 3 in Madow (33) and the results of Cramer[1] (6) yield the following:

i) $(\hat{B},\hat{R})$ are distributed asymptotically as a bivariate normal with mean vector (B_N, R_N) and covariance matrix given by

$$\begin{bmatrix} \bar{V}(\hat{B}) & \overline{Cov}(\hat{B},\hat{R}) \\ \overline{Cov}(\hat{B},\hat{R}) & \bar{V}(\hat{R}) \end{bmatrix} .$$

ii) When $B_N=R_N$,

$$\frac{\hat{B}-\hat{R}}{\sqrt{\hat{V}(\hat{B}-\hat{R})}}$$

is distributed asymptotically as a normal random variable with mean zero and variance 1 where

$$\hat{V}(\hat{B}-\hat{R}) = \hat{V}(\hat{B}) + \hat{V}(\hat{R}) - 2\,\widehat{Cov}(\hat{B},\hat{R}) \quad ,$$

$$\hat{V}(\hat{B}) = \frac{1}{\left(\sum\limits^{n} \frac{x_i^2}{n}\right)^2} \frac{1}{n}\left(1-\frac{n}{N}\right) \sum^{n} \frac{(y_i-\hat{B}_N x_i)^2 x_i^2}{n-1} ,$$

---

[1] Cramer, H. (6, pp. 364-367).

$$\hat{\bar{V}}(\hat{R}) = \frac{1}{\bar{x}_n^2} \frac{1}{n}(1 - \frac{n}{N}) \sum_{i=1}^{n} \frac{(y_i - \hat{R}_N x_i)^2}{n - 1} \quad \text{and}$$

$$\hat{\overline{Cov}}(\hat{B},\hat{R}) = \frac{1}{\sum \frac{x_i^2}{n} \bar{x}_n} \frac{1}{n}(1 - \frac{n}{N}) \sum^{n} \frac{(y_i - \hat{B}_N x_i)(y_i - \hat{R}_N x_i) x_i}{n - 1}$$

iii) from ii), when $B_N = R_N$

$$\frac{(\hat{B}-\hat{R})^2}{\bar{V}(\hat{B}-\hat{R})}$$

is distributed asymptotically as a Chi-square random variable with one degree of freedom (d.f.). This result follows from a theorem in Fisz[1] (11).

iv) Finally, since $(\hat{B},\hat{R})$ are distributed asymptotically as a bivariate normal, $((\hat{B}-\hat{R}),\hat{R})$ is distributed asymptotically as a bivariate normal with mean vector (B_N-R_N,R_N) and covariance matrix

$$\begin{bmatrix} \bar{V}(\hat{B}-\hat{R}) & \overline{Cov}(\hat{B}-\hat{R},\hat{R}) \\ \overline{Cov}(\hat{B}-\hat{R},\hat{R}) & \bar{V}(\hat{R}) \end{bmatrix}$$

where

$$\overline{Cov}(\hat{B}-\hat{R}) = \overline{Cov}(\hat{B},\hat{R}) - \bar{V}(\hat{R}) \quad .$$

Now consider the mixture estimator

---

[1]Fisz, Marek (11, pp. 183-184).

$$\hat{\theta}_1 = \hat{\alpha}\hat{B}\bar{X}_N + (1-\hat{\alpha})\hat{R}\bar{X}_N = \hat{\alpha}(\hat{B}-\hat{R})\bar{X}_N + \hat{R}\bar{X}_N \tag{IV.15}$$

where

$$\hat{\alpha} = \frac{1}{1+\hat{\delta}^2} \quad \text{and}$$

$$\hat{\delta}^2 = \max\left\{0, \frac{(\hat{B}-\hat{R})^2}{\hat{\hat{V}}(\hat{B}-\hat{R})} - 1\right\} .$$

In the following, we derive properties of $\hat{\theta}_1$ in a large sample situation. In one instance we concentrate on the limiting distributions of sample statistics and in another instance we utilize order in probability concepts. The analysis is divided into two cases. The first is when

i) $\lim\limits_{N\to\infty} B_N-R_N = 0$ and the second,

ii) $\lim\limits_{N\to\infty} B_N-R_N = B-R \neq 0$.

From Equation IV.14

$$\hat{B}-\hat{R} = (B_N-R_N) + O_p\left(\frac{1}{\sqrt{n}}\right)$$

and we assume

$$\hat{B}-\hat{R} = (B-R) + O_p\left(\frac{1}{\sqrt{n}}\right) .$$

In practice, however, we are faced with a single population. Hence, once we have derived expressions that involve B, for example, we shall replace it by B_N.

Consider now the first case wherein

$$\lim_{N\to\infty} (B_N - R_N) = 0 \quad .$$

Let

i) G_n denote the event that $|\hat{R}-R_N| \geq \varepsilon/2$

ii) H_n denote the event that $|\hat{\alpha}(\hat{B}-\hat{R})| \geq \varepsilon/2$ and

iii) I_n denote the event that $|\hat{B}-\hat{R}| \geq \varepsilon/2$.

We know that $\hat{R}$ and $\hat{B}-\hat{R}$ are consistent for R_N and 0, respectively. Given $\varepsilon/2 > 0$ and $\delta/2 > 0$, there exists N_ε such that $\forall n \geq N_\varepsilon$,

$$P[G_n] < \delta/2 \qquad \text{and} \qquad P_r[I_n] < \delta/2$$

Now, given $\varepsilon > 0$ and $\delta > 0$, we have

$$P[\,|\hat{\theta}_1/\bar{X}_N - R_N| \geq \varepsilon] = P[\,|\hat{\alpha}(\hat{B}-\hat{R}) + \hat{R}-R_N| \geq \varepsilon] \leq P[H_n] + P[G_n] \quad .$$

The random variable $\hat{\alpha}$ is bounded by 0 and 1. Hence it follows that $P[H_n] \leq P[I_n]$. We now have

$$P[\,|\hat{\theta}_1/\bar{X}_N - R_N| \geq \varepsilon] \leq P[G_n] + P[I_n] < \delta \qquad \forall \quad n \geq N_\varepsilon \ .$$

Hence when $\lim(B_N - R_N) = 0$, plim $\hat{\theta}_1 = \bar{Y}$.

On the other hand, when

$$\lim_{N\to\infty} B_N - R_N = B - R \neq 0 \quad ,$$

$$\hat{\alpha} = \frac{1}{1 + \max\left\{0, \dfrac{(B_N - R_N)^2 + 2(B_N - R_N)O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{n})}{\hat{V}(\hat{B} - \hat{R})} - 1\right\}}$$

$$= \frac{1}{1 + \max\left\{0, \dfrac{(B - R)^2 + 2(B - R)O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{n})}{\hat{V}(\hat{B} - \hat{R})} - 1\right\}}$$

$$= O_p(\tfrac{1}{n}) \quad .$$

We then have

$$\text{plim } \hat{\theta}_1 = \text{plim } O_p(\tfrac{1}{n}) \text{ plim } (\hat{B} - \hat{R})\bar{X}_N + \text{plim } \hat{R}\bar{X}_N = \bar{Y} \quad .$$

The above results follow by applications of Theorems I and II in Appendix 8. Hence $\hat{\theta}_1$ is consistent for $\bar{Y}$.

When $\lim(B_N - R_N) = 0$, the following calculations show that under the limiting distribution of $\hat{\theta}_1$, $\overline{A\text{-}V}(\hat{\theta}_1) \leq \overline{A\text{-}V}(\hat{R}\bar{X}_N)$.

An equivalent expression of $\hat{\alpha}$ is given by

$$\hat{\alpha} = 1 \quad \text{if} \quad \chi' \leq 1$$

$$= \frac{1}{\chi'}, \quad \text{if} \quad \chi' > 1$$

where

$$\chi' = (\hat{B}-\hat{R})^2/\hat{\bar{V}}(\hat{B}-\hat{R}) \ .$$

The limiting distribution of $\hat{\alpha}$ when

$$\lim_{N\to\infty} B_N - R_N = 0$$

is the distribution of

$$\hat{\alpha} = 1 \quad \text{if} \quad 0 < \chi \leq 1$$
$$= \frac{1}{\chi} \quad \text{if} \quad \chi > 1$$

where χ denotes a Chi-squared random variable with one degree of freedom. This result follows from Fisz[1].

Using the asymptotic distributions of the components of $\hat{\theta}_1$ we now calculate the large sample M.S.E. of $\hat{\theta}_1$. Recall that these asymptotic distributions were given immediately following the Assumption (Condition W).

$$(\hat{\theta}_1 - \bar{Y})^2 = [\hat{\alpha}(\hat{B}-\hat{R}) + \hat{R} - R]^2\bar{X}^2$$

$$= \bar{X}^2[\hat{\alpha}^2(\hat{B}-\hat{R})^2 + 2(\hat{B}-\hat{R})(\hat{R}-R)\hat{\alpha} + (\hat{R}-R)^2] \ . \tag{IV.16}$$

Now,

$$\bar{E}[\hat{\alpha}^2(\hat{B}-\hat{R})^2] = \int_0^1 \bar{V}(\hat{B}-\hat{R})xf(x)dx + \int_1^\infty \frac{\bar{V}(\hat{B}-\hat{R})}{x}f(x)dx \tag{IV.17}$$

---

[1]Fisz, Marek (11, pp. 183-184).

where f(x) denotes the Chi-square density with one degree of freedom. For the following integration we use Pearson's Tables of the Incomplete Gamma Function (44).

The first term in Equation IV.17 is equal to

$$\bar{V}(\hat{B}-\hat{R})\int_0^1 \frac{x^{1/2}e^{-x/2}}{\sqrt{2\pi}}dx = .0969\ \bar{V}(\hat{B}-\hat{R}) \quad . \tag{IV.18}$$

In the following, we shall simplify $\bar{V}(\hat{B}-\hat{R})$ to $\bar{V}$. Similarly, $\overline{\text{Cov}}(\hat{B}-\hat{R})$ will be denoted $\overline{\text{Cov}}$. The second term in Equation IV.17 is

$$\bar{V}\int_1^\infty \frac{x^{-3/2}e^{-x/2}}{\sqrt{2\pi}}dx = \frac{\bar{V}}{\sqrt{2\pi}}[2e^{-1/2} - \{\int_0^\infty (2u)^{-1/2}e^{-u}2du$$

$$- \int_0^{1/2} (2u)^{-1/2}e^{-u}2du\}]$$

The above equation follows by integration by parts and a change of variables. Evaluating the integral yields

$$\bar{V}\int_1^\infty \frac{x^{-3/2}e^{-x/2}}{\sqrt{2\pi}}dx = .2013\ \bar{V} \quad . \tag{IV.19}$$

Combining equations IV.18 and IV.19 we have

$$\bar{X}^2\bar{E}[\hat{\alpha}^2(\hat{B}-\hat{R})^2] = .2982\ \bar{X}^2\bar{V} \quad .$$

The second term in Equation IV.16 is

$$\bar{X}^2 2[\hat{\alpha}(\hat{B}-\hat{R})(\hat{R}-R)] \quad .$$

We now express $\hat{\alpha}$ in the more convenient form,

$$\hat{\alpha} = 1 \qquad \text{if} \qquad |\hat{B}-\hat{R}| \leq \sqrt{\bar{\bar{V}}}$$

$$= \frac{1}{\chi} \qquad \text{if} \qquad |\hat{B}-\hat{R}| > \sqrt{\bar{\bar{V}}} \quad .$$

We then have,

$$\bar{E}[\hat{\alpha}(\hat{B}-\hat{R})(\hat{R}-R)] = \int_{-\sqrt{\bar{\bar{V}}}}^{\sqrt{\bar{\bar{V}}}} \int_{-\infty}^{\infty} (\hat{B}-\hat{R})(\hat{R}-R)g(\hat{B}-\hat{R},\hat{R})d(\hat{B}-\hat{R})d(\hat{R})$$

$$+ \bar{V} \int_{\sqrt{\bar{\bar{V}}}}^{\infty} \int_{-\infty}^{\infty} (\hat{B}-\hat{R})^{-1}(\hat{R}-R)g(\hat{B}-\hat{R},\hat{R})d(\hat{B}-\hat{R})d(\hat{R})$$

$$+ \bar{V} \int_{-\infty}^{-\sqrt{\bar{\bar{V}}}} \int_{-\infty}^{\infty} (\hat{B}-\hat{R})^{-1}(\hat{R}-R)g(\hat{B}-\hat{R},\hat{R})d(\hat{B}-\hat{R})d(\hat{R})$$

(IV.20)

The symbol $g(\hat{B}-\hat{R},\hat{R})$ denotes the bivariate normal density of the random variable $(\hat{B}-\hat{R},\hat{R})$. From multivariate normal theory we know that the conditional distribution of $\hat{R}$ given $(\hat{B}-\hat{R})$ is normal with mean

$$R + \frac{\overline{\text{Cov}}}{\bar{\bar{V}}}(\hat{B}-\hat{R}) \quad .$$

Hence for each of the three integrals in Equation IV.20 we have

$$\int_{-\infty}^{\infty} (\hat{R}-R)h(\hat{R}\,|\hat{B}-\hat{R})d\hat{R} = \frac{\overline{\text{Cov}}(\hat{B}-\hat{R})}{\bar{\bar{V}}}$$

where $h(\hat{R}|\hat{B}-\hat{R})$ denotes the conditional normal density function of $\hat{R}$ given $(\hat{B}-\hat{R})$. Equation IV.20 then reduces to

$$\bar{E}[\hat{\alpha}(\hat{B}-\hat{R})(\hat{R}-R)] = 2\int_{0}^{\sqrt{\bar{V}}} (\hat{B}-\hat{R})^2 \frac{\overline{Cov}}{\bar{V}} k(\hat{B}-\hat{R})d(\hat{B}-\hat{R})$$

$$+ 2\int_{\sqrt{\bar{V}}}^{\infty} \overline{Cov}\; k(\hat{B}-\hat{R})d(\hat{B}-\hat{R}) \quad . \qquad \text{(IV.21)}$$

The symbol $k(\hat{B}-\hat{R})$ denotes the marginal normal density of $(\hat{B}-\hat{R})$. In Equation IV.21 we have utilized the symmetry of the normal distribution in combining integrals. The first term in Equation IV.21 is

$$2\int_{0}^{\sqrt{\bar{V}}} \frac{\overline{Cov}}{\bar{V}^{3/2}} \frac{(\hat{B}-\hat{R})^2}{\sqrt{2\pi}} \exp[-\frac{(\hat{B}-\hat{R})^2}{2\bar{V}}]d(\hat{B}-\hat{R}) = .1939\ \overline{Cov}.$$

The second term in Equation IV.21 is

$$2\int_{\sqrt{\bar{V}}}^{\infty} \overline{Cov}\; g(\hat{B}-\hat{R})d(\hat{B}-\hat{R}) = .3198\ \overline{Cov} \quad .$$

The above two results yield

$$\bar{X}_N^2\, 2\bar{E}[\hat{\alpha}(\hat{B}-\hat{R})(\hat{R}-R)] = (1.027)\bar{X}_N^2\,\overline{Cov} \quad .$$

Finally, using the definitions of $\bar{V}$ and $\overline{Cov}$,

$$\bar{E}[\hat{\alpha}^2(\hat{B}-\hat{R})^2] + 2\bar{E}[\hat{\alpha}(\hat{B}-\hat{R})(\hat{R}-R)]$$

$$= .2982[\bar{V}(\hat{B}) + \bar{V}(\hat{R}) - 2\overline{Cov}(\hat{B},\hat{R})] + 1.027[\overline{Cov}(\hat{B},\hat{R}) - \bar{V}(\hat{R})]$$

$$\doteq .30\ \bar{V}(\hat{B}) - .73\ \bar{V}(\hat{R}) + .43\ \overline{Cov}(\hat{B},\hat{R}) \qquad \text{(IV.22)}$$

Under Model VI with $\mathcal{V}_i = \sigma^2 v_i$, and omitting terms of order $1/n^2$, we have

$$\bar{E}[\hat{\alpha}^2(\hat{B}-\hat{R})^2 + 2\hat{\alpha}(\hat{B}-\hat{R})(\hat{R}-R)]$$

$$= \frac{\sigma^2}{n}(1-\frac{n}{N})[\frac{.30}{(\sum^N \frac{x_i^2}{N})^2}\sum^N \frac{x_i^2}{N-1} - \frac{.73}{\bar{X}_N^2} + \frac{.43}{\sum \frac{x_i^2}{N}}]$$

$$\doteq \frac{\sigma^2}{n}(1-\frac{n}{N})[\frac{.73}{\sum \frac{x_i^2}{N}} - \frac{.73}{\bar{X}_N^2}]$$

$$= -\frac{\sigma^2}{n}(1-\frac{n}{N})(.73)[\sum\frac{N(x_i-\bar{X}_N)^2}{N-1}][\frac{1}{\bar{X}_N^2 \sum^N \frac{x_i^2}{N}}] . \qquad \text{(IV.23)}$$

From Equations IV.23 and IV.16 we have that

$$\overline{AV}(\hat{R}\bar{X}_N) - \overline{AV}(\hat{\theta}_1) = .73\frac{\sigma^2}{n}(1-\frac{n}{N})[\sum\frac{N(x_i-\bar{X}_N)^2}{N-1}][\frac{1}{\sum^N \frac{x_i^2}{N}}] . \qquad \text{(IV.24)}$$

Hence $\overline{A\text{-}V}(\hat{\theta}_1) \le \overline{A\text{-}V}(\hat{R}\bar{X}_N)$.

We now consider the M.S.E. of $\hat{\theta}_1$ when

$$\lim_{N\to\infty} (B_N - R_N) \neq 0 \quad .$$

From Equation IV.15,

$$(\hat{\theta}_1 - \bar{Y}) = [\hat{\alpha}(\hat{B}-\hat{R}) + \hat{R}-R]\bar{X}_N$$

and

$$(\hat{\theta}_1-\bar{Y})^2 = [\hat{\alpha}^2(\hat{B}-\hat{R})^2 + 2\hat{\alpha}(\hat{B}-\hat{R})(\hat{R}-R) + (\hat{R}-R)^2]\bar{X}_N{}^2 \quad . \qquad \text{(IV.25)}$$

We have already shown that $\hat{\alpha} = O_p(\frac{1}{n})$. Omitting expectation of terms of order in probability greater than $1/n$ yields

$$\bar{E}[(\hat{\theta}_1 - \bar{Y})^2] = \frac{1}{n}(1 - \frac{n}{N}) \sum^{N} \frac{(y_i - R_N x_i)^2}{N-1} \quad .$$

Let us summarize the results of Section 1 in Theorem F.3.

Let the model be

$$\mathcal{E}(y_i|x_i) = \beta x_i$$

$$\mathcal{C}ov(y_i,y_j|x_i,x_j) = \sigma^2 \qquad i = j$$

$$= 0 \qquad i \neq j$$

where β and σ^2 are unknown. Given that we sample n tuples (y,x) with SRS w/o replacement, the mixture estimator

$$\hat{\theta}_1 = \hat{\alpha}\hat{B}\bar{X}_N + (1 - \hat{\alpha})\hat{R}\bar{X}_N \ ,$$

with

$$\hat{\alpha} = \frac{1}{1+\hat{\delta}^2} \quad , \quad \hat{\delta}^2 = \max\left\{o, \frac{(\hat{B}-\hat{R})^2}{\hat{V}(\hat{B}-\hat{R})} - 1\right\}$$

is

i) consistent for the population mean and

ii) $\overline{A\text{-}V}(\hat{\theta}_1) \le \overline{A\text{-}V}(\hat{R}\bar{X}_N)$.

2. Mixture type estimator 2

Consider the model

$$\mathcal{E}(y_i | x_i) = \beta x_i$$

$$\mathcal{C}ov(y_i, y_j | x_i, x_j) = \sigma^2 \qquad i = j$$
$$= 0 \qquad i \neq j$$

where β is unknown and σ^2 is known. As in the previous section, assume that we select a sample of n tuples (y_i, x_i) by SRS w/o replacement and that $\bar{X}_N$ is known prior to sample selection. Furthermore, consider the estimators

$$\hat{B}\bar{X}_N = \frac{\sum^n y_i x_i}{\sum^n x_i^2} \bar{X}_N$$

and

$$\hat{R}\bar{X}_N = \frac{\sum^n y_i}{\sum^n x_i} \bar{X}_N \quad .$$

Let $\hat{\theta} = \alpha\hat{B}\bar{X}_N + (1-\alpha)\hat{R}\bar{X}_N$. If we minimize

$$\mathcal{V}(\hat{\theta}|s) + \frac{1}{\sqrt{n}}\alpha^2(\hat{B}-\hat{R})^2\bar{X}_N^{\,2}$$

with respect to α, we obtain

$$\hat{\alpha} = 1/[1 + \frac{(\hat{B}-\hat{R})^2}{\sqrt{n}\,\sigma^2}[\frac{n\bar{x}_n^{\,2}\sum^n x_i^{\,2}}{\sum^n (x_i - \bar{x}_n)^2}]] \qquad \text{(IV.26)}$$

and the resulting mixture estimator

$$\hat{\theta}_2 = \hat{\alpha}\hat{B}\bar{X}_N + (1-\hat{\alpha})\hat{R}\bar{X}_N$$

is consistent for $\bar{Y}_N$. We state in the following theorem, some relevant properties of $\hat{\theta}_2$.

Theorem F.4:

Given the sequence of populations, $\langle pop \rangle_{n=1}^{\infty}$, in section 1. Given that we sample n tuples (y,x) with SRS and w/o replacement. The estimator

$$\hat{\theta}_2 = \hat{\alpha}\hat{B}\bar{X}_N + (1-\hat{\alpha})\hat{R}\bar{X}_N$$

where $\hat{\alpha}$ is defined in Equation IV.26 is a consistent estimator of the population mean and has a $\overline{\text{M.S.E.}}$ of

$$\text{i)}\ \overline{\text{M.S.E.}}(\hat{\theta}_2) = \frac{\bar{X}_N^{\,2}}{(\Sigma\frac{x_i^{\,2}}{N})^2}(1-\frac{n}{N})\frac{1}{n}\sum^N \frac{(y_i - B_N x_i)^2 x_i^{\,2}}{N-1}$$

when $\lim B_N - R_N = 0$ and

ii) $$\overline{\text{M.S.E.}}(\hat{\theta}_2) = \frac{1}{n}\left(1-\frac{n}{N}\right)\sum^{N}\frac{(y_i-R_Nx_i)^2}{N-1}$$

$$+\frac{\bar{X}_N^{\,2}}{n}(B_N-R_N)^2\left[\frac{1}{\frac{1}{\sqrt{n}}\;\frac{(B_N-R_N)^2K}{\sigma^2}}\right]^2$$

when $\lim B_N - R_N \neq 0$, where

$$K = \bar{X}_N^{\,2}\sum^{N}\frac{x_i^{\,2}}{N}\Big/\sum^{N}\frac{(x_i-\bar{X}_N)^2}{N-1}$$

and where expectation of terms of order in probability larger than $1/n$ are omitted.

Proof:

Utilizing the sequence of populations, $\langle pop_n \rangle$, as defined in section 1, we have

$$\hat{B}-\hat{R} = B_N-R_N + \frac{\sum^{n} x_iu_i}{\sum^{n} x_i^{\,2}} - \frac{\sum^{n} s_i}{\sum^{n} x_i}$$

$$= B-R + O_p\left(\frac{1}{\sqrt{n}}\right) \quad .$$

The order in probability is established by Theorem II in Appendix 8 and the consistency of sample means with respect to the sequence of populations. As in section 1 we shall examine the estimator $\hat{\theta}_2$ in the two limiting cases.

i) $\lim B_N - R_N = B - R = 0$ and

ii) $\lim B_N - R_N = B - R \neq 0.$

Hence, in the analysis we shall replace B_N, R_N, etc. by their limiting values B, R, etc. Now

$$(\hat{B}-\hat{R})^2 = (B-R)^2 + 2(B-R)O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{n}) \quad .$$

From Equation IV.26, we write

$$\hat{\alpha} = 1/[1 + \sqrt{n}\ \frac{(\hat{B}-\hat{R})^2}{\sigma^2}\{K + O_p(\frac{1}{\sqrt{n}})\}] \quad .$$

where the order in probability follows as a result of Theorem II in Appendix 8 and where

$$K = \bar{X}_N^{\,2} \sum^N \frac{x_i^{\,2}}{N} \Big/ \sum^N \frac{(x_i - \bar{X}_N)^2}{N-1} \quad .$$

When B = R,

$$\hat{\alpha} = 1/[1 + \frac{O_p(\frac{1}{\sqrt{n}})}{\sigma^2}\{K + O_p(\frac{1}{\sqrt{n}})\}] \quad .$$

Hence plim $\hat{\alpha} = 1.$

When $B \neq R$,

$$\hat{\alpha} = 1/[1 + \frac{\sqrt{n}}{\sigma^2}[(B-R)^2 + O_p(\frac{1}{\sqrt{n}})][(K + O_p(\frac{1}{\sqrt{n}}))] \quad .$$

Hence plim $\hat{\alpha} = 0.$ The

$$\text{plim } \hat{\theta}_2 = \text{plim } \hat{\alpha} \text{ plim } \hat{B}\bar{X}_N + \text{plim } (1-\hat{\alpha}) \text{ plim } \hat{R}\bar{X}_N$$

$$= \bar{Y} \quad \text{if} \quad B = R$$

$$= \bar{Y} \quad \text{if} \quad B \neq R.$$

Hence $\hat{\theta}_2$ is consistent for $\bar{Y}$. We now find the $\overline{\text{M.S.E.}}(\hat{\theta}_2)$.

$$\hat{\theta}_2 - \bar{Y}_N = (\hat{\alpha}\hat{B}\bar{X}_N + (1-\hat{\alpha})\hat{R}\bar{X}_N - \bar{Y}_N \quad .$$

Replacing the population characteristics with their limiting values,

$$\hat{\theta}_2 - \bar{Y} = \bar{X}[\hat{\alpha}(B-R) + \hat{\alpha} \frac{\sum^n x_i u_i}{\sum x_i^2} + (1-\hat{\alpha}) \frac{\sum^n s_i}{\sum x_i}] \quad .$$

When B = R, we have

i) $\hat{\alpha} = \dfrac{1}{1 + O_p(\frac{1}{\sqrt{n}})} = 1 + O_p(\frac{1}{\sqrt{n}})$

ii) $1 - \hat{\alpha} = O_p(\frac{1}{\sqrt{n}})$

iii) $\hat{\alpha}^2 = 1 + O_p(\frac{1}{\sqrt{n}})$

iv) $(1-\hat{\alpha})^2 = O_p(\frac{1}{n})$

v) $\hat{\alpha}^2 \left(\dfrac{\sum^n x_i u_i}{\sum^n x_i^2}\right)^2 = O_p(\frac{1}{n})$

vi) $(1-\hat{\alpha})^2\left(\frac{\sum^n s_i}{\sum^n x_i}\right)^2 = O_p\left(\frac{1}{n^2}\right)$

vii) $(1-\hat{\alpha})\hat{\alpha}\,\frac{\sum^n x_i u_i}{\sum x_i^2}\;\frac{\sum^n s_i}{\sum^n x_i} = O_p\left(\frac{1}{n^{3/2}}\right)$.

Now,

$$(\hat{\theta}_2-\bar{Y})^2 = \bar{X}^2\left[\hat{\alpha}^2\left(\frac{\sum^n x_i u_i}{\sum^n x_i^2}\right)^2 + (1-\hat{\alpha})^2\left(\frac{\sum^n s_i}{\sum^n x_i}\right)^2\right.$$

$$\left. + 2\hat{\alpha}(1-\hat{\alpha})\,\frac{\sum x_i u_i}{\sum x_i^2}\;\frac{\sum s_i}{\sum x_i}\right]$$

Omitting terms of order in probability greater than 1/n,

$$\bar{E}[(\hat{\theta}_2-\bar{Y})^2] = \bar{X}^2\;\bar{E}\left[\hat{\alpha}^2\left(\frac{\sum^n x_i u_i}{\sum x_i^2}\right)^2\right] .$$

Replacing the limiting population values by their corresponding values, we have

$$\bar{E}[\hat{\theta}_2-\bar{Y}_N)^2] = \frac{\bar{X}_N^{\,2}}{\left(\sum^N \frac{x_i^{\,2}}{N}\right)^2}\;\frac{1}{n}\left(1-\frac{n}{N}\right)\sum^N \frac{(y_i-B_N x_i)^2 x_i^{\,2}}{N-1} . \qquad \text{(IV.27)}$$

When $B \neq R$, we have

i) $\hat{\alpha} = O_p(\frac{1}{\sqrt{n}})$

ii) $\hat{\alpha}^2 = O_p(\frac{1}{n})$

iii) $1 - \hat{\alpha} = 1 + O_p(\frac{1}{\sqrt{n}})$

iv) $(1 - \hat{\alpha})^2 = 1 + O_p(\frac{1}{\sqrt{n}})$.

Also

$$\hat{\theta}_2 - \bar{Y} = \bar{X}[\hat{\alpha}(B-R) + \{\hat{\alpha} \sum^n \frac{x_i u_i}{\Sigma x_i^2} + (1 - \hat{\alpha}) \frac{\sum^n s_i}{\sum^n x_i}\}]$$

Omitting expectation of terms of order in probability larger than 1/n,

$$\bar{E}(\hat{\theta}_2 - \bar{Y})^2 = \bar{X}^2(B-R)^2\bar{E}(\hat{\alpha}^2) + \bar{X}^2\bar{E}[(\frac{\sum^n s_i}{\sum x_i})^2]$$

$$+ 2\bar{X}^2(B-R)\bar{E}[\hat{\alpha} \frac{\sum^n s_i}{\sum x_i}] \quad .$$

The

$$\bar{E}[\hat{\alpha}^2] = \frac{1}{n}/(\frac{1}{\sqrt{n}} + C)^2 \qquad \text{(IV.28)}$$

where

$$C = \frac{(B-R)^2}{\sigma^2} K \ .$$

Equation IV.28 follows as a result of a Taylor's series expansion omitting expectation of terms of order in probability larger than 1/n.

Similarly, it can be shown that

$$\bar{E}[\hat{\alpha}\frac{\sum^{n} s_i}{\sum x_i}] = 0$$

omitting expectation of terms of order in probability larger than 1/n.

Thus when $B \neq R$,

$$\bar{E}(\hat{\theta}_2 - \bar{Y}_N)^2 = \bar{X}_N{}^2(B_N-R_N)^2\bar{E}[\hat{\alpha}^2] + \frac{1}{n}(1-\frac{n}{N}) \sum^{N} \frac{(y_i-R_Nx_i)^2}{N-1} \quad .(IV.29)$$

where

$$\bar{E}[\hat{\alpha}^2] = \frac{1}{n}/[\frac{1}{\sqrt{n}} + \frac{(B-R)^2}{\sigma^2}K]^2 \quad .Q.E.D.$$

Let us summarize the results of Theorem F.4. The estimator $\hat{\theta}_2$ is consistent for the population mean. When $B = R$, the $\overline{M.S.E.}(\hat{\theta}_2)$ is equal to the $\overline{M.S.E.}(\hat{B}\bar{X}_N)$. Incidentally, since the biases of $\hat{B}\bar{X}_N$, $\hat{R}\bar{X}_N$ and $\hat{\theta}_2$ are of order 1/n, to our degree of approximation, we may replace $\overline{M.S.E.}$ by $\bar{V}$.

When $B \neq R$,

$$\bar{V}(\hat{\theta}_2) - \bar{V}(\hat{R}\bar{X}_N) = \frac{1}{n}\bar{X}_N{}^2(B_N-R_N)^2[\frac{1}{\frac{1}{\sqrt{n}} + \frac{(B_N-R_N)^2}{\sigma^2}K}]^2 \quad .$$

This difference decreases as $(B_N-R_N)^2$ increases or as n increases. However, if B-R is quite small, this difference will be large.

We note here that if σ^2 is not known we may use

$$\hat{\sigma}^2 = \sum^n \frac{(y_i-\hat{B}x_i)^2}{n-1}$$

in $\hat{\alpha}$ as an estimate of σ^2. This modification does not alter the consistency of $\hat{\theta}_2$. The properties of $\hat{\theta}_2$ when $\hat{\sigma}^2$ is used are discussed in Appendix 7.

If we wish the large sample variance of $\hat{\theta}_2$ to be that given in Equation IV.27 when $B = R$ and to be the large sample variance of $\hat{R}\bar{X}_N$ otherwise, and to the same degree of approximation, i.e., omitting expectation of terms of order in probability greater than 1/n, we need only to modify $\hat{\alpha}$ to the form

$$\hat{\alpha}' = 1/[1 + n^{\frac{1}{2}+\varepsilon} \quad \frac{(\hat{B}-\hat{R})^2}{\sigma^2}\{\frac{\bar{x}_n^{\,2} \sum^n \frac{x_i^2}{n}}{\sum \frac{(x_i-\bar{x}_n)^2}{n}}\}] \qquad \text{(IV.30)}$$

where $0 < \varepsilon < 1/2$. The discussion of $\hat{\alpha}'$ is given in Appendix 7.

We note that although

$$\mathcal{E}(\hat{R}\bar{X}_N|s) = \mathcal{E}(\hat{B}\bar{X}_N|s) = \beta\bar{X}_N,$$

$\mathcal{E}(\hat{\theta}_2|s)$ need not necessarily be $\beta\bar{X}_N$. Of course if $(\hat{B}-\hat{R})$ is uncorrelated with $\hat{B}\bar{X}_N$ and $\hat{R}\bar{X}_N$ under the superpopulation model,

$$\mathcal{E}(\hat{\theta}_2|s) = \beta\bar{X}_N .$$

3. Mixture type estimator 3

In this section we consider an estimator that is consistent for the population mean and has smaller anticipated variance than the usual ratio estimator, $\hat{R}\bar{X}_N$.

Consider the model,

$$\mathcal{E}(y_i|x_i) = \beta x_i$$

$$\mathcal{C}ov(y_i,y_j|x_i,x_j) = \sigma^2 \qquad i = j$$
$$= 0 \qquad i \neq j$$

where β and σ^2 are unknown and $\bar{X}_N$ is known. We select a sample of n tuples (y, x) with SRS and w/o replacement. We shall discuss the estimator

$$\hat{\gamma} = \bar{y}_n + \hat{B}(\bar{X}_N - \bar{x}_n)$$

where

$$\hat{B} = \frac{\sum^{n} x_i y_i}{\sum x_i^2} .$$

Utilizing the sequence of populations constructed in section 1, it is easy to show that $\hat{\gamma}$ is a consistent estimator of $\bar{Y}_N$. In addition,

$$\mathcal{E}(\hat{\gamma}|s) = \beta\bar{x}_n + \beta(\bar{X}_N - \bar{x}_n) = \beta\bar{X}_N = \mathcal{E}(\bar{Y}_N)$$

and hence $\hat{\gamma}$ is unbiased for $\beta\bar{X}_N$ under the model.

Defining

$$B_N = \frac{\sum^{N} x_i y_i}{\sum^{N} x_i^2}$$

and $u_i = y_i - B_N x_i$, it follows that $\sum^{N} u_i x_i = 0$. The $\overline{\text{M.S.E.}}(\hat{\gamma})$ is

$$\begin{aligned} \text{M.S.E.}(\hat{\gamma}) &= E[(\hat{\gamma} - \bar{Y}_N)^2] \\ &= E[(\bar{U}_n - \bar{U}_N - (\hat{B} - B_N)(\bar{x}_n - \bar{X}_N))^2] \\ &= \bar{E}[(\bar{U}_n - \bar{U}_N)^2] , \end{aligned}$$

where we have omitted expectation of terms of order in probability larger than 1/n. Finally,

$$\overline{\text{M.S.E.}}(\hat{\gamma}) = \frac{1}{n}(1 - \frac{n}{N}) \sum_{i=1}^{N} \frac{(y_i - \bar{Y}_N - B_N(x_i - \bar{X}_N))^2}{N - 1} .$$

Now since $\hat{R}\bar{X}_N$ is also consistent for $\bar{Y}_N$, unbiased under the model, and possesses a $\overline{\text{M.S.E.}}$ of the same order, we compare $\hat{R}\bar{X}_N$ and $\hat{\gamma}$ on the basis of A-V.

First, consider $\mathcal{V}(\hat{R}\bar{X}_N - \bar{Y}_N | s)$ and $\mathcal{V}(\hat{\gamma} - \bar{Y}_N | s)$. We have

$$\mathcal{V}(\hat{R}\bar{X}_N - \bar{Y}_N | s) = \frac{\bar{X}_N^2 \sigma^2}{n\bar{x}_n^2} + \frac{\sigma^2}{N} - \frac{2\bar{X}_N \sigma^2}{n\bar{x}_n}$$

and

$$\mathcal{V}(\hat{Y}-\bar{Y}_N|s) = \frac{\sigma^2}{n} + \frac{(\bar{X}_N-\bar{x}_n)^2}{\sum\limits^{n} x_i^2}\sigma^2 - \frac{2(\bar{X}_N-\bar{x}_n)\bar{x}_n\sigma^2}{\sum\limits^{n} x_i^2}$$

$$+ \frac{\sigma^2}{N} - \frac{2(\bar{X}_N-\bar{x}_n)n\bar{x}_n\sigma^2}{N\sum\limits^{n} x_i^2} - \frac{2\sigma^2}{N} .$$

Upon completing squares, we have

$$\mathcal{V}(\hat{Y}-\bar{Y}_N|s) = \sigma^2[\frac{1}{n} - \frac{1}{N} + \frac{N\bar{X}_N^{\ 2}-N\bar{x}_n^{\ 2}-2n\bar{x}_n\bar{X}_N+2n\bar{x}_n^{\ 2}}{N\sum\limits^{n} x_i^2}] \qquad \text{(IV.31)}$$

Also $\mathcal{V}(\hat{R}\bar{X}_N - \bar{Y}_N|s)$ may be written as

$$\mathcal{V}(\hat{R}\bar{X}_N - \bar{Y}_N|s) = \sigma^2[\frac{1}{n} - \frac{1}{N} + \frac{N\bar{X}_N^{\ 2}-N\bar{x}_n^{\ 2}-2n\bar{x}_n\bar{X}_N+2n\bar{x}_n^{\ 2}}{nN\bar{x}_n^{\ 2}}]. \text{(IV.32)}$$

Taking the difference of Equations IV.31 and IV.32 we have

$$\mathcal{V}(\hat{Y} - \bar{Y}_N|s) - \mathcal{V}(\hat{R}\bar{X}_N - \bar{Y}_N|s) = -\frac{K}{N}\,\frac{\sum\limits^{n}(x_i-\bar{x}_n)^2}{n\sum\limits^{n} x_i^2\bar{x}_n^{\ 2}}$$

where

$$K = (\bar{X}_N - \frac{n}{N}\bar{x}_n)^2 - (\bar{x}_n - \frac{n}{N}\bar{x}_n)^2 .$$

Hence, if $K > 0$, i.e., if $|\bar{X}_N - \frac{n}{N}\bar{x}_n| > (1-\frac{n}{N})|\bar{x}_n|$, $\mathcal{V}(\hat{Y} - \bar{Y}_N|s)$ is smaller than $\mathcal{V}(\hat{R}\bar{X}_N - \bar{Y}_N|s)$ and conversely.

This suggests an estimator $\hat{\theta}_3$ defined by

$$\hat{\theta}_3 = \hat{R}\bar{X}_N \quad \text{if} \quad |\bar{X}_N - \frac{n}{N}\bar{x}_n| \leq |\bar{x}_n|(1 - \frac{n}{N})$$

$$= \hat{\gamma} \quad \text{otherwise.}$$

$\hat{R}\bar{X}_N$ and $\hat{\gamma}$ are both conditionally unbiased under the model. $\hat{\theta}_3$ will have smaller anticipated variance than either $\hat{R}\bar{X}_N$ or $\hat{\gamma}$ since conditional on the sample $\hat{\theta}_3$ is the preferred estimator for every sample. In addition, we now show that $\hat{\theta}_3$ is a consistent estimator of $\bar{Y}_N$.

Let

i) B denote the event that $K \geq 0$

ii) B^c denote the event that $K < 0$

iii) $\cap$ denote the intersection of two events.

Given $\varepsilon > 0$,

$$P[|\hat{\theta}_3 - \bar{Y}_N| > \varepsilon] = P[|\hat{\theta}_3 - \bar{Y}_N| > \varepsilon | B]P[B]$$

$$+ P[|\hat{\theta}_3 - \bar{Y}_N| > \varepsilon | B^c]P[B^c]$$

where $P[A|B]$ denotes the conditional probability of event A given that event B has occurred. We then have

$$P[|\hat{\theta}_3 - \bar{Y}_N| > \varepsilon] = P[|\hat{\theta}_3 - \bar{Y}_N| > \varepsilon \cap B] + P[|\hat{\theta}_3 - \bar{Y}_N| > \varepsilon \cap B^c]$$

$$\leq P[|\hat{\gamma} - \bar{Y}_N| > \varepsilon] + P[|\hat{R}\bar{X}_N - \bar{Y}_N| > \varepsilon] \quad .$$

Since $\hat{\gamma}$ and $\hat{R}\bar{X}_N$ are both consistent for $\bar{Y}_N$, it follows that $\hat{\theta}_3$ is consistent for $\bar{Y}_N$.

In summary, in sections 1, 2 and 3 we investigate three consistent mixture type estimators of $\bar{Y}_N$. Each mixture type estimator is shown to have at most as large $\overline{A\text{-}V}$ as the usual ration estimator $\hat{R}\bar{X}_N$.

V. REGRESSION MODELS

In Chapter V we consider two types of regression models.

Def. A model of the form

$$\mathcal{E}(y_i|x_i) = g(U_i)\alpha + \beta f(x_i)$$

$$\mathcal{C}ov(y_i,y_j|x_i,x_j) = \mathcal{V}_i \qquad i = j$$

$$= 0 \qquad i \neq j$$

where α and β are unknown will be termed a regression model. The function g may be a function of any characteristic of unit U_i except y_i.

A. Regression Model 1

Consider a regression model of the form

$$\mathcal{E}(y_i|x_i) = \alpha + \beta(x_i - \bar{X}_N)$$

$$\mathcal{C}ov(y_i,y_j|x_i,x_j) = \mathcal{V}_i \qquad i = j$$

$$= 0 \qquad i \neq j$$

where α and β are unknown. Assume that $\bar{X}_N$ is known. In the following, we shall assume a without replacement design of size n with inclusion probabilities π_i. Each sample of size n consists of n tuples (y,x).

Given a sample of size n, Regression Model 1 may be written in the vector form,

$$y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = XB + e \qquad \text{where}$$

$$X = \begin{bmatrix} 1 & x_1 - \bar{X}_N \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n - \bar{X}_N \end{bmatrix} \qquad \text{is a } n \times 2 \text{ matrix,}$$

y and e are n x 1 vectors and

$$B = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \qquad \text{is a } 2 \times 1 \text{ vector.}$$

Given the sample of size n, e is distributed with mean zero and covariance matrix $\mathcal{V}$ where

$$\mathcal{V} = \begin{bmatrix} \mathcal{V}_1 & & 0 \\ & \ddots & \\ 0 & & \mathcal{V}_n \end{bmatrix}.$$

Given the sample of size n, an estimator of α may be obtained from

$$\hat{B} = (X'\gamma^{-1}X)^{-1} X'\gamma^{-1}y$$

and is given by

$$\hat{\alpha}' = \frac{1}{\Delta'}[\sum^{n} \frac{(x_i-\bar{X}_N)^2}{nN\gamma_i} \sum^{n} \frac{y_i}{nN\gamma_i} - \sum^{n} \frac{(y_i(x_i-\bar{X}_N)}{Nn\gamma_i} \sum^{n} \frac{(x_i-\bar{X}_N)}{nN\gamma_i}]$$

(V.1)

where

$$\Delta' = \sum^{n} \frac{1}{nN\gamma_i} \sum^{n} \frac{(x_i-\bar{X}_N)^2}{nN\gamma_i} - (\sum^{n} \frac{(x_i-\bar{X}_N)}{nN\gamma_i})^2 \quad .$$

We assume that $\Delta' \neq 0$ for all samples s. Under the model,

$$\bar{Y}_N = \alpha + \bar{e}_N \quad ,$$

so a possible estimator of $\bar{Y}_N$ is $\hat{\alpha}'$ given in Equation V.1. Under the model, $\hat{\alpha}'$ is the least squares estimator of α. If we replace $n\gamma_i$ by π_i, where $0 < \pi_i < 1$,

$$\sum^{N} \pi_i = n \quad ,$$

we have

$$\hat{\alpha} = \frac{1}{\Delta} [\sum^{n} \frac{(x_i-\bar{X}_N)^2}{N\pi_i} \sum^{n} \frac{y_i}{N\pi_i} - \sum^{n} \frac{y_i(x_i-\bar{X}_N)}{N\pi_i} \sum^{n} \frac{(x_i-\bar{X}_N)}{N\pi_i}]$$

(V.2)

where

$$\Delta = \sum^{n} \frac{(x_i - \bar{X}_N)^2}{N\pi_i} \sum^{n} \frac{1}{N\pi_i} - \left(\sum^{n} \frac{(x_i - \bar{X}_N)}{N\pi_i}\right)^2 .$$

The estimator $\hat{\alpha}$ has the following properties which we state as a theorem.

Theorem V.1:

Given a without replacement design of size n of tuples (y,x) and given that $\bar{X}_N$ is known. Assume that Regression Model 1 holds. Then, the estimator $\hat{\alpha}$ given in Equation V.2

i) is consistent for $\bar{Y}_N$ for any set of inclusion probabilities $\{\pi_i\}$,

ii) is unbiased for α under the model and

iii) possesses an $\overline{AV}$ that is minimized for π_i proportional to $\sqrt{\gamma_i}$

where terms of order in probability larger than 1/n are omitted.

Proof:

We first prove ii) of Theorem V.1.

$$\mathcal{E}(\hat{\alpha}\,|\,s) = \mathcal{E}\Big[\frac{1}{\Delta}\Big\{\sum^{n} \frac{(x_i - \bar{X}_N)^2}{N\pi_i} \sum^{n} \frac{(\alpha + \beta(x_i - \bar{X}_N) + e_i)}{N\pi_i}$$

$$- \sum^{n} \frac{(x_i - \bar{X}_N)}{N\pi_i}(\alpha + \beta(x_i - \bar{X}_N) + e_i) \sum^{n} \frac{(x_i - \bar{X}_N)}{N\pi_i}\Big\}\Big]$$

$$= \frac{1}{\Delta}\Big[\alpha \sum^{n} \frac{(x_i - \bar{X}_N)^2}{N\pi_i} \sum^{n} \frac{1}{N\pi_i} - \alpha\left(\sum^{n} \frac{(x_i - \bar{X}_N)}{N\pi_i}\right)^2\Big] = \alpha .$$

Proof of i). In Chapter IV we constructed a sequence of populations such that the H.T. estimator

$$\sum^{n} \frac{z_i}{N\pi_i}$$

minus its expectation over the sampling design was of order in probability $\frac{1}{\sqrt{n}}$. We shall assume the same construction here. Since,

$$\sum^{n} \frac{(x_i-\bar{X}_N)^2}{N\pi_i}, \quad \sum^{n} \frac{y_i}{N\pi_i}, \quad \sum^{n} \frac{y_i(x_i-\bar{X}_N)}{N\pi_i} \quad \text{and} \quad \sum^{n} \frac{x_i-\bar{X}_N}{N\pi_i}$$

are all H.T. estimators, using Theorem I in Appendix 8,

$$\text{plim } \hat{\alpha} = \bar{Y}_N \ .$$

Proof of iii). The

$$\text{A-V}(\hat{\alpha}) = E\mathcal{V}(\hat{\alpha} - \bar{Y}_N | s) \quad .$$

Omitting terms of order in probability larger than $\frac{1}{n}$,

$$\text{A-V}(\hat{\alpha}) = \sum^{N} \frac{\mathcal{V}_i}{N^2\pi_i} - \sum_{i=1}^{N} \frac{\mathcal{V}_i}{N^2} \ . \qquad \text{(V.3)}$$

Using the same minimization techniques in Section B, π_i proportional to $\sqrt{\mathcal{V}_i}$ minimizes the $\overline{\text{A-V}}(\hat{\alpha})$ given in Equation V.3. Q.E.D.

We now give the M.S.E. of $\hat{\alpha}$ omitting terms of order in probability larger than 1/n. Define

$$B^* = \sum^{N} (y_i-\bar{Y}_N)(x_i-\bar{X}_N)/\sum^{N}(x_i-\bar{X}_N)^2$$

and

$$y_i = \bar{Y}_N + B^*(x_i - \bar{X}_N) + u_i \quad . \qquad (V.4)$$

It follows that

$$\sum^{N} u_i = 0$$

and

$$\sum^{N} u_i(x_i - \bar{X}_N) = 0 \quad .$$

The M.S.E. $(\hat{\alpha}) = E[(\hat{\alpha} - \bar{Y}_N)^2]$. From Equation V.4,

$$E[(\hat{\alpha} - \bar{Y}_N)^2] = E[\frac{1}{\Delta^2}(\sum^{n} \frac{(x_i - \bar{X}_N)^2}{N\pi_i} \sum^{n} \frac{u_i}{N\pi_i} - \sum^{n} \frac{x_i - \bar{X}_N}{N\pi_i} \sum^{n} \frac{u_i(x_i - \bar{X}_N)}{N\pi_i})^2].$$

Omitting terms of order in probability larger than $1/n$ and using a Taylor's series expansion of $1/\Delta^2$,

$$\bar{E}(\hat{\alpha} - \bar{Y}_N)^2 = \bar{E}[(\sum^{n} \frac{u_i}{N\pi_i})^2]$$

$$= \frac{1}{N^2} \sum_{i \neq j}^{N} (\pi_i \pi_j - \pi_{ij})(\frac{u_i}{\pi_i} - \frac{u_j}{\pi_j})^2$$

where $u_i = y_i - \bar{Y}_N - B^*(x_i - \bar{X}_N)$ and π_{ij} denotes the joint inclusion probability of units i and j.

B. Regression Model 2

Recall that the regression estimator in simple random sampling is

$$y_n + \hat{\beta}(\bar{X}_N - \bar{x}_n)$$

where

$$\hat{\beta} = \frac{\sum^{n} y_i (x_i - \bar{x}_n)}{\sum (x_i - \bar{x}_n)^2} \ .$$

Basically, the regression estimator is of the form mean of the y_i + product of the regression coefficient and $(\bar{X}_N - \bar{x}_n)$. A possible extension of the regression estimator to the unequal probability sampling case is

$$L_1 = \sum^{n} \frac{y_i}{Nnp_i} + \hat{\beta}' (\bar{X}_N - \sum^{n} \frac{x_i}{Nnp_i})$$

where

i) $0 < p_i < 1$, $\sum^{N} p_i = 1$

ii) $\hat{\beta}' = \frac{1}{D} \sum^{n} (\frac{x_i}{Np_i} - \frac{1}{n} \sum^{n} \frac{x_i}{Np_i}) \frac{y_i}{Np_i}$

iii) $D = \sum^{n} (\frac{x_i}{Np_i} - \frac{1}{n} \sum^{n} \frac{x_i}{Np_i})^2$.

We assume that $D \neq 0 \ \forall s$.

Under Regression Model 1,

$$\mathcal{E}(L_1 | s) = \sum^{n} \frac{\alpha}{Nnp_i} + \beta \sum^{n} \frac{x_i - \bar{X}_N}{Nnp_i} + (\bar{X}_N - \sum^{n} \frac{x_i}{Nnp_i}) [\frac{\alpha \sum^{n} (\frac{x_i}{Np_i} - \frac{1}{n} \sum \frac{x_i}{Np_i}) \frac{1}{Np_i}}{D}$$

$$+ \frac{\beta}{D} \sum^{n} (\frac{x_i}{Np_i} - \frac{1}{n} \sum \frac{x_i}{Np_i}) (\frac{x_i - \bar{X}_N}{Np_i})] \ .$$

Since this expectation is not equal to $\alpha = \mathcal{E}(\bar{Y}_N)$, L_1 is biased under the model. Hence, L_1 does not 'zero out' the first component of the A-V as given in Equation III.2.

We now consider another regression model, denoted Regression Model 2 for which L_1 appears appropriate.

The following model is called Regression Model 2.

Let

$$\mathcal{E}(y_i | x_i) = \alpha\sqrt{\gamma_i} + \beta x_i$$

$$\mathcal{C}ov(y_i, y_j | x_i, x_j) = \gamma_i \qquad i = j$$
$$= 0 \qquad i \neq j$$

where α and β are unknown. As in Regression Model 1 we wish to estimate $\bar{Y}_N$ by selecting a sample of n tuples (y,x) without replacement. Assume that $\bar{X}_N$ is known.

Under Regression Model 2, one can show that the estimator L_1 is the least squares estimator of

$$\mathcal{E}(\bar{Y}_N) = \alpha \sum^{N} \frac{\sqrt{\gamma_i}}{N} + \beta\bar{X}_N \quad .$$

Furthermore,

$$\mathcal{E}(L_1 | s) = \mathcal{E}[\sum^{n} \frac{y_i}{Nnp_i} + \beta'(\bar{X}_N - \frac{1}{n} \sum^{n} \frac{x_i}{Np_i}) | s]$$

$$= \sum^{N} \alpha \frac{\sqrt{\gamma_i} + \beta x_i}{Nnp_i} + \beta(\bar{X}_N - \sum^{n} \frac{x_i}{Nnp_i})$$

$$= \alpha \sum^{N} \frac{\sqrt{\gamma_i}}{N} + \beta\bar{X}_N \quad .$$

We now consider other properties of L_1. The

$$\text{A-V}(L_1) = E\mathcal{V}(L_1 - \bar{Y}_N | s) \quad .$$

Omitting terms of order in probability larger than 1/n,

$$\overline{\text{A-V}}(L_1) = \sum^{N} \frac{\mathcal{V}_i}{N^2 n p_i} - \sum^{N} \frac{\mathcal{V}_i}{N^2} \quad .$$

As in previous sections, p_i proportional to $\sqrt{\mathcal{V}_i}$ minimizes the above quantity. Define

$$F = \sum^{N} \left(\frac{y_i}{Np_i} - \bar{Y}_N\right)\left(\frac{x_i}{Np_i} - \bar{X}_N\right) p_i \Big/ \sum^{N} \left(\frac{x_i}{Np_i} - \bar{X}_N\right)^2 p_i$$

and define

$$\frac{y_i}{Np_i} = \bar{Y}_N + F\left(\frac{x_i}{Np_i} - \bar{X}_N\right) + \frac{e_i}{Np_i} \quad .$$

It follows that

$$\sum^{N} \frac{e_i}{Np_i}\left(\frac{x_i}{Np_i} - \bar{X}_N\right) p_i = 0$$

and

$$\sum^{N} \frac{e_i}{N} = 0 \quad .$$

Now,

$$L_1 = \bar{Y}_N + (\hat{\beta}' - F)\left(\bar{X}_N - \sum^{n} \frac{x_i}{nNp_i}\right) + \sum^{n} \frac{e_i}{nNp_i} \quad .$$

Also,

$$(\hat{\beta}'-F) = \sum^{n}\left(\frac{x_i}{Np_i} - \frac{1}{n}\sum\frac{x_i}{Np_i}\right)\frac{e_i}{Np_i} \Big/ \sum^{n}\left(\frac{x_i}{Np_i} - \frac{1}{n}\sum\frac{x_i}{Np_i}\right)^2 = 0_p\left(\frac{1}{\sqrt{n}}\right)$$

by Theorem II in Appendix 8. Hence,

$$E[(L_1 - \bar{Y}_N)^2] = E[(\hat{\beta}'-F)^2(\bar{X}_N - \sum^{n}\frac{x_i}{Nnp_i})^2 + (\sum^{n}\frac{e_i}{Nnp_i})^2$$

$$+ 2(\hat{\beta}'-F)(\bar{X}_N - \sum^{n}\frac{x_i}{Nnp_i})(\sum^{n}\frac{e_i}{Nnp_i})] \quad .$$

Omitting terms of order in probability larger than 1/n, we have

$$\overline{\text{M.S.E.}}(L_1) = \bar{E}[(\sum^{n}\frac{e_i}{Nnp_i})^2]$$

$$= \frac{1}{N^2}\sum_{i\neq j}(\pi_i\pi_j - \pi_{ij})\left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j}\right)^2$$

where

$$e_i = y_i - \bar{Y}_N - F\left(\frac{x_i}{Np_i} - \bar{X}_N\right) \quad ,$$

$$\pi_i = np_i$$

and π_{ij} denotes the joint inclusion probability of units i and j.

To summarize the above results concerning L_1, we have

Theorem V.2:

Given a without replacement design of size n of tuples (y,x) and given that $\bar{X}_N$ is known. Assume that Regression

Model 2 holds. Then the estimator L_1 has the following properties:

i) L_1 is unbiased under the model.

ii) L_1 is the least squares estimator of $\mathscr{E}(\bar{Y}_N)$ under the model.

iii) Omitting terms of order in probability larger than 1/n,

a) the $\overline{\text{A-V}}(L_1)$ is minimized for p_i proportional to $\sqrt{\gamma_i}$ and

b) the $\overline{\text{M.S.E.}}(L_1)$ is given by

$$\frac{1}{N^2}\sum_{i\neq j}^{N}(\pi_i\pi_j-\pi_{ij})\left(\frac{e_i}{\pi_i}-\frac{e_j}{\pi_j}\right)^2$$

where

$$e_i = y_i - \bar{Y}_N - F\left(\frac{x_i}{Np_i} - \bar{X}_N\right)$$

iv) L_1 is consistent for $\bar{Y}_N$.

In summary, we have considered two regression type models in Chapter V. In both cases we have presented consistent estimators of $\bar{Y}_N$ and found the 'best' inclusion probabilities with respect to the $\overline{\text{A-V}}$ omitting terms of order in probability larger than 1/n.

VI. ADDITIONAL RESULTS CONCERNING SURVEY DESIGNS

In this chapter we concern ourselves with some miscellaneous topics including variable sample size sampling designs, cost functions and stratified replacement sampling.

A. Variable Sample Size Designs

Godambe and Joshi (17) in their Theorem 7.1 prove that for any given fixed sample design p with inclusion probabilities π_i, it is impossible to construct a variable size sample design p^* with the same inclusion probabilities π_i, and such that the variance of the H.T. estimator is smaller under p^* than the fixed sample design p for all possible values of y. As stated by Hanurav (22), this theorem implies that fixed sample size designs using the H.T. estimator are admissible in the class of sampling designs with common inclusion probabilities but does not completely exclude variable sampling designs. In Theorem IV.1 we give a partial justification for fixed sample size designs with respect to a certain class of variable sample size designs.

Theorem VI.1:

For any variable sample size design (denoted by G) with inclusion probability $\pi_i > 0 \ \forall_i$ and joint inclusion probability $\pi^G_{ij} = \pi_i \pi_j$ there exists a fixed sample size without replacement design (denoted by F) with sample size

$n = \sum_{i=1}^{N} \pi_i$ (n, integer) and inclusion probability π_{ij}^{F}

using the same $\pi_i > 0$ and such that $V_G(\text{H.T.}) \geq V_F(\text{H.T.})$ whenever $y_i \geq 0 \ \forall_i$.

Proof:

Under the sampling design G,

$$V_G(\text{H.T.}) = V_G\left(\sum_{i \epsilon s} \frac{y_i}{N\pi_i}\right) = \frac{1}{N^2} \sum_{i=1}^{N} \left(\frac{y_i}{\pi_i}\right)^2 \pi_i(1-\pi_i)$$

$$+ \frac{2}{N^2} \sum_{i}^{N} \sum_{j>i} \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} (\pi_{ij}^{G} - \pi_i \pi_j) \quad .$$

Under the sampling design F,

$$V_F(\text{H.T.}) = V_F\left(\sum_{i=1}^{n} \frac{y_i}{N\pi_i}\right) = \frac{1}{N^2}\left[\sum_{i=1}^{N} \frac{y_i^{\,2}}{\pi_i} - \sum_{i=1}^{N} y_i^{\,2}\right]$$

$$+ \frac{1}{N^2} \sum_{i \neq j}^{N} \left(\frac{\pi_{ij}^{F}}{\pi_i \pi_j} - 1\right) y_i y_j \quad .$$

Hence,

$$V_G(\text{H.T.}) - V_F(\text{H.T.}) = -\frac{1}{N^2}\left[\sum_{i \neq j}^{N} \left\{\frac{\pi_{ij}^{F}}{\pi_i \pi_j} - 1\right\} y_i y_j\right] \qquad \text{(VI.1)}$$

A sufficient condition that the right hand side of Equation VI.I. be nonnegative is that $\pi_{ij}^{F}/\pi_i\pi_j \leq 1$ and $y_i \geq 0 \ \forall_i$. We note that whenever $y_i \geq 0 \ \forall_i$, any unequal probability fixed sample size design with $\pi_i > 0 \ \forall_i$ and possessing a nonnegative unbiased estimator of variance will possess the property $\pi_{ij}^F/\pi_i\pi_j \leq 1 \; \forall i \neq j$.

An example of a design G is Poisson sampling with $\pi_i = np_i$ where

$$\sum^{N} p_i = 1 \quad ,$$

and n, integer. An example of a design F is Sampford's (51) rejective sampling scheme of size n that maintains the inclusion probability $\pi_i = np_i$. We shall show by way of a counter example, the necessity of $y_i \geq 0 \; \forall_i$. Consider the artificial population of N = 4, n = 2 and suppose Sampford's scheme is used. Let $y_1 = y_2 = 0$ and $y_3 = -15$, $y_4 = 1$. Then, substituting into Equation VI.1 we have

$$V_G(\text{H.T.}) - V_F(\text{H.T.}) = \frac{1}{N^2}\left(1 - \frac{\pi_{34}^F}{\pi_3\pi_4}\right) (-15) < 0 \quad .$$

B. Cost Function

In Chapter IV we neglected consideration of the cost of the survey design. We implicitly assumed that the cost of observing a unit in the population was the same as any other unit. Hence, by specifying a fixed sample size we have implicitly fixed cost. Hájek (21) has stated sufficient conditions that will insure the optimality (minimum A-V) of a survey design over the class T_5 of estimators for a fixed expected cost of the form

$$\sum_{i=1}^{N} c_i \pi_i \quad .$$

Recall that the difference estimator d_1 in Chapter IV satisfied the lower bound result of Appendix 5 when π_i is proportional to $\sqrt{\gamma_i}$. We now generalize the problem somewhat by minimizing the A-V(d_1) with respect to a fixed expected cost

$$C_o = \sum_{i=1}^{N} c_i \pi_i \quad .$$

From Chapter IV, the

$$\text{A-V}(d_1) = \frac{1}{N^2} \sum_{i=1}^{N} \frac{\gamma_i}{\pi_i} - \frac{1}{N^2} \sum_{i=1}^{N} \gamma_i \quad .$$

We now minimize

$$\sum^{N} \frac{\gamma_i}{\pi_i}$$

under the condition that

$$C_o = \sum_{i=1}^{N} c_i \pi_i \quad .$$

Recall that a special form of Hölder's inequality is

$$\sum^{N} s_i t_i \leq (\sum^{N} s_i^2)^{1/2} (\sum^{N} t_i^2)^{1/2}$$

where $s_i, t_i > 0 \quad \forall_i$.

Letting $t_i = \sqrt{c_i \pi_i}$ and $s_i = \sqrt{\gamma_i / \pi_i}$ we have

$$\frac{1}{C_o}\left(\sum^N \sqrt{\gamma_i c_i}\right)^2 \leq \sum^N \frac{\gamma_i}{\pi_i} . \qquad \text{(VI.2)}$$

If we set

$$\pi_i^* = \frac{C_o\sqrt{\gamma_i / c_i}}{\sum^N \sqrt{\gamma_i c_i}} ,$$

the R.H.S. of Equation VI.2 is equal to the L.H.S. Therefore π_i^* gives the minimum value of

$$\sum^N \gamma_i / \pi_i$$

under the condition of fixed expected cost, C_o. It is assumed that $\pi_i^* \leq 1$, $\forall_i$. We state the above results as Theorem VI.2.

Theorem VI.2:

Given Model II in Chapter IV. Given also that

i) we have a specified expected cost

$$C_o = \sum_{i=1}^N c_i \pi_i$$

where c_i is the cost of sampling unit i and

ii) $$0 < \frac{C_o\sqrt{\gamma_i / c_i}}{\sum^N \sqrt{\gamma_i c_i}} < 1 \quad \forall_i .$$

Then, the

$$A\text{-}V(d_1) = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\mathcal{V}_i}{\pi_i} - \frac{1}{N^2}\sum_{i=1}^{N}\mathcal{V}_i$$

is minimized for

$$\pi_i^* = \frac{C_o\sqrt{\mathcal{V}_i/c_i}}{\sum^N \sqrt{\mathcal{V}_i c_i}} \quad v_i \quad .$$

We note that since $\mathcal{V}_i$, c_i and C_o are ordinarily fixed in advance,

$$\sum_{i=1}^{N} \pi_i^*$$

need not sum to n. Hence, Theorem VI.2 will ordinarily specify a variable sample size design.

C. Stratification in With Replacement Sampling

Let N be the size of a finite population, let p_m, m=1,...,N denote the selection probabilities where

$$\sum_{m=1}^{N} p_m = 1 \quad ,$$

and let n denote the sample size for a fixed sample size design.

Assume that it is possible to group the N units into L strata such that

$$\sum_{j=1}^{N_i} np_{(i)j} = n_i$$

where N_i denotes the size of the i^{th} stratum, $p_{(i)j}$ denotes the original selection probability of unit j in the i^{th} stratum,

$$\sum_{i=1}^{L} \sum_{j=1}^{N_i} p_{(i)j} = 1 = \sum_{m=1}^{N} p_m$$

and n_i is an integer,

$$\sum_{i=1}^{L} n_i = n.$$

Let A denote the unequal probability with replacement sampling scheme of fixed size n with selection probabilities p_m,m=1,...,N and the estimator

$$a = \frac{1}{N} \sum_{m=1}^{n} \frac{y_m}{np_m} \quad .$$

Selection of the n units to be achieved by selecting unit i with probability p_i at each of n draws.

Let B denote the unequal probability with replacement sampling scheme when the population is stratified into L strata with sampling performed independently within each stratum with selection probability

$$p'_{(i)j} = \frac{n}{n_i} p_{(i)j}$$

for the j^{th} unit in the i^{th} stratum and the stratified estimator

$$b = \sum_{i=1}^{L} \frac{N_i}{N}\left(\frac{1}{n_i} \sum_{j=1}^{n_i} \frac{y_{(i)j}}{N_i p'_{(i)j}}\right) \quad .$$

We note that estimator b reduces to estimator a.

We now show that $V_B(a) \leq V_A(a)$. We know that

$$V_A(a) = \frac{1}{nN^2} \sum_{i=1}^{N} p_i \left(\frac{y_i}{p_i} - Y\right)^2$$

and

$$V_B(a) = \sum_{i=1}^{L} \left(\frac{N_i}{N}\right)^2 \frac{1}{N_i{}^2 n_i} \sum_{j=1}^{N_i} p'_{(i)j} \left(\frac{y_{(i)j}}{p'_{(i)j}} - Y_i\right)^2$$

$$= \frac{1}{N^2} \sum_{i=1}^{L} \frac{1}{n_i} \sum_{j=1}^{N_i} \frac{n}{n_i} p_{(i)j} \left(\frac{y_{i(j)}}{\frac{n}{n_i} p_{(i)j}} - \frac{\frac{n}{n_i}}{\frac{n}{n_i}} Y_i\right)^2$$

$$= \frac{1}{nN^2} \sum_{i=1}^{L} \sum_{j=1}^{N_i} p_{(i)j} \left(\frac{y_{(i)j}}{p_{(i)j}} - \frac{n}{n_i} Y_i\right)^2$$

where

$$Y_i = \sum_{j=1}^{N_i} y_{(i)j} \quad .$$

The $V_A(a)$ can be written as

$$\frac{1}{nN^2} \sum_{i=1}^{L} \sum_{j=1}^{N_i} \left[p_{(i)j} \left(\frac{y_{(i)j}}{p_{(i)j}}\right)^2 - 2y_{(i)j} Y + p_{(i)j} Y^2\right]$$

$$= \frac{1}{nN^2} \sum^{L} \sum^{N_i} p_{(i)j} \left(\frac{y_{(i)j}}{p_{(i)j}}\right)^2 - \frac{1}{nN^2} Y^2$$

and hence

$$V_A(a) - V_B(a) = \frac{1}{nN^2} \sum_{i=1} \frac{2n}{n_i} Y_i^2 - \frac{1}{nN^2} \sum^{L} \frac{n}{n_i} Y_i^2 - \frac{1}{nN^2} Y^2$$

$$= \frac{1}{nN^2}\left[\sum_{i=1}^{L} \frac{n}{n_i} Y_i^2 - (\sum^{L} Y_i)^2\right]$$

$$= \frac{1}{nN^2} \sum^{L} \frac{n_i}{n}\left(\frac{Y_i}{n_i/n} - Y\right)^2 \geq 0 \quad .$$

Hence, $V_A(a) \geq V_B(a)$. The usefulness of this result is illustrated in the following theorem.

Theorem VI.3:

If in the preceding development $n_i = 2\ V_i$, then $V_A(a) \geq V_B(a) \geq V_C(c)$ where C denotes stratified w/o replacement sampling of 2 units per stratum with

i) inclusion probabilities

$$\pi_{i(j)} = 2p'_{(i)j}$$

for unit j in the i^{th} stratum

ii) the w/o replacement sampling design of size two within stratum i with the estimator

$$\frac{1}{N_i}\left(\frac{Y_{(i)j}}{2p'_{(i)j}} + \frac{Y_{(i)k}}{2p'_{(i)k}}\right)$$

is superior to the with replacement sampling scheme of size two using the selection probabilities $p'_{i(j)}$ and the usual mean, and

iii) c is the estimator

$$\sum_{i=1}^{n/2} \frac{N_i}{N} \frac{1}{2}\left(\frac{y_{(i)j}}{N_i p'_{(i)j}} + \frac{y_{(i)k}}{N_i p'_{(i)k}}\right)$$

Proof:

When $n_i = 2$ for $i=1,\ldots,L$

$$V_A(a) - V_B(a) = \frac{1}{nN^2}\left[\left(\frac{n}{2} - 1\right) \sum_{i=1}^{n/2} Y_i^2 - \sum_{i \neq j}^{n/2} Y_i Y_j\right]$$

$$= \frac{1}{nN^2}\left[\frac{1}{2} \sum_{i \neq j}^{n/2} (Y_i - Y_j)^2\right] \geq 0 \quad .$$

Using Durbin's (9) scheme for selection of the two units per stratum where the first unit is selected with $p'_{(i)j}$ and the second unit with

$$p'_{(i)j|k} = \frac{p'_{(i)k}}{1 + A_i}\left[\frac{1}{1 - 2p'_{(i)j}} + \frac{1}{1 - 2p'_{(i)k}}\right]$$

where

$$A_i = \sum_{j=1}^{N_i} \frac{p'_{(i)j}}{1 - 2p'_{(i)j}} \quad ,$$

we have that $V_A(a) \geq V_B(a) \geq V_C(c)$.

D. Without Versus With Replacement Sampling

Raj (46) gave the following theorem concerning with and without replacement sampling of fixed size n.

Theorem VI.4:

Let

i) p_i denote the selection probabilities of a with replacement design of size n

ii) $\pi_i = np_i$ denote the inclusion probability of unit i in w/o replacement sampling

iii) π_{ij} denote the inclusion probability of units i and j in w/o replacement sampling of size n.

Then, a sufficient condition for the variance of the H.T. estimator,

$$\sum_{i=1}^{n} \frac{y_i}{\pi_i} \quad ,$$

to have smaller variance than the with replacement estimator

$$\sum_{i=1}^{n} \frac{y_i}{np_i}$$

is that

$$\pi_{ij} > \frac{n-1}{n} \pi_i \pi_j \quad \forall_{i \neq j} \quad .$$

A necessary condition for without replacement sampling of size n to be superior to with replacement sampling under the conditions of Raj is given by the following.

Theorem VI.5 (Narain, 40):

A necessary condition for the without replacement estimator

$$\sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

to have smaller variance than the with replacement estimator when $\pi_i = np_i$ is that

$$\pi_{ij} \leq \frac{2(n-1)}{n} \pi_i \pi_j \quad .$$

We give another theorem concerning with versus without replacement sampling designs of size n.

Theorem VI.6:

Given a without replacement design of fixed size n such that $\pi_i = np_i$ and $\pi_{ij} = n(n-1)p_i p_j + \Delta_{ij}$. If

$$\sum_{i \neq j}^{N} \Delta_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \geq 0$$

then,

$$\mathrm{Var}_{w/o}(\mathrm{H.T.}) \leq V_{w/}\left(\frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{np_i}\right)$$

where $V_{w/o}$(H.T.) denotes the variance of the H.T. estimator under the without replacement design and $V_{w/}$ denotes variance under the with replacement design with selection probabilities p_i.

Proof:

Using the Yates and Grundy (56) form of the variance we have

$$V_{w/o}(\text{H.T.}) = \frac{1}{N^2} \sum_{i \neq j} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 .$$

Since $\pi_{ij} = n(n-1)p_i p_j + \Delta_{ij}$,

$$V_{w/o}(\text{H.T.}) = \frac{1}{N^2} \sum_{i \neq j} \frac{1}{n} p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 - \frac{1}{N^2} \sum_{i \neq j} \Delta_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

$$= \frac{1}{N^2 n} \sum_{i=1}^{N} p_i \left(\frac{y_i}{p_i} - Y\right)^2 - \frac{1}{N^2} \sum_{i \neq j} \Delta_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 .$$

Since

$$\frac{1}{N^2} \sum_{i \neq j} \Delta_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \geq 0$$

we are done.

An example of a design such that $\pi_{ij} = n(n-1)p_i p_j + \Delta_{ij}$ and

$$\sum_{i \neq j} \Delta_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \geq 0$$

is obtainable under the following situation. Group the units into L strata such that

$$\sum_{i=1}^{N_i} \pi_i = 2 , \quad \sum_{i=1}^{L} N_i = N$$

and select samples of size 2 within each stratum using Durbin's (9) scheme or Fuller's (14) scheme B.

In both of the above schemes we have that

$$\sum_{i \neq j}^{N} \Delta_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 \geq 0 \quad .$$

We extend the above theorem to consider without replacement designs with variable sample size.

Theorem VI.7:

Let n denote the expected sample size of a variable sample size design. Let the without replacement sample size design of variable sample size be such that $\pi_i = np_i < 1$ and $\pi_{ij} = n(n-1)p_i p_j + \Delta_{ij}$. If the expression

$$\frac{n-1}{n} \sum_{i=1}^{N} y_i^2 - \frac{1}{n} \sum_{i \neq j}^{N} \Delta_{ij} \frac{y_i y_j}{p_i p_j} \geq 0 \quad ,$$

then

$$V_{w/o}(\text{H.T.}) \leq V_{w/}\left(\frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{Nnp_i}\right)$$

where $V_{w/o}$ and $V_{w/}$ are the same as in Theorem VI.6.

Proof:

Using the Horwitz and Thompson (28) form of the variance of

$$\frac{1}{N} \sum_{i \varepsilon s} \frac{y_i}{\pi_i} \quad ,$$

and substituting $\pi_{ij} = n(n-1)p_i p_j + \Delta_{ij}$ we have

$$V_{w/o}(\text{H.T.}) = \frac{1}{nN^2}\left(\sum^N \frac{y_i^2}{p_i} - Y^2\right) + \frac{1}{N^2}\left(\frac{Y^2}{n} - Y^2\right)$$

$$+ \frac{1}{N^2} \sum_{i \neq j} \left(\frac{n(n-1)p_i p_j + \Delta_{ij}}{n^2 p_i p_j}\right) y_i y_j$$

$$= \frac{1}{nN^2}\left(\sum^N \frac{y_i^2}{p_i} - Y^2\right) + \frac{1}{N^2 n} \sum^N y_i^2 - \frac{1}{N^2} \sum^N y_i^2$$

$$+ \frac{1}{N^2} \sum_{i \neq j}^{N} \frac{\Delta_{ij} y_i y_j}{n^2 p_i p_j}$$

$$= \frac{1}{N^2 n}\left(\sum^N \frac{y_i^2}{p_i} - Y^2\right) - \frac{1}{N^2}\left(\frac{n-1}{n}\right) \sum y_i^2 + \frac{1}{nN^2} \sum_{i \neq j}^{N} \frac{\Delta_{ij} y_i y_j}{n^2 p_i p_j} .$$

Hence if

$$\left(\frac{n-1}{n}\right) \sum^N y_i^2 - \frac{1}{n} \sum_{i \neq j} \Delta_{ij} \frac{y_i y_j}{p_i p_j} > 0 ,$$

$$V_{w/}\left(\sum_{i=1}^{n} \frac{y_i}{Nnp_i}\right) - V_{w/o}(\text{H.T.}) \geq 0 .$$

An example of when the conditions of the above theorem are satisfied is Poisson Sampling with inclusion probability $\pi_i = np_i$ and $\bar{Y}_N = 0$.

E. Unbiased Variance Estimation

The class of estimators, T_5, appears to encompass most of the estimators in practice. For completeness, we give an unbiased estimator of the variance of unbiased estimators of $\bar{Y}_N$ in the class T_5 under sampling designs admitting unbiased estimators of variance.

Since

$$V(t_5) = E[(\sum_{i\varepsilon s} \beta_{is} y_i)^2] - \bar{Y}_N^2$$

an unbiased estimator of $V(t_5)$ is

$$\hat{V}(t_5) = t_5^2 - \text{unbiased est.}(\bar{Y}_N^2)$$

$$= (\sum_{i\varepsilon s} \beta_{is} y_i)^2 - \text{unbiased est.}[\frac{1}{N^2}(\sum_{i=1}^{N} y_i^2 + \sum_{i\neq j}^{N} y_i y_j)]$$

$$= (\sum_{i\varepsilon s} \beta_{is} y_i)^2 - \frac{1}{N^2}[\sum_{i\varepsilon s} \frac{y_i^2}{\pi_i} + \sum_{(i\neq j)\varepsilon s} \frac{y_i y_j}{\pi_{ij}}]$$

$$= \sum_{i\varepsilon s} \beta_{is}^2 y_i^2 + \sum_{(i\neq j)\varepsilon s} \beta_{is}\beta_{js} y_i y_j$$

$$- \frac{1}{N^2}[\sum_{i\varepsilon s} \frac{y_i^2}{\pi_i} + \sum_{(i\neq j)\varepsilon s} \frac{y_i y_j}{\pi_{ij}}]$$

$$= \sum_{i\varepsilon s} (\beta_{is}^2 - \frac{1}{N^2\pi_i}) y_i^2 + \sum_{(i\neq j)\varepsilon s} (\beta_{is}\beta_{js} - \frac{1}{N^2\pi_{ij}}) y_i y_j \quad .$$

If

$$\beta_{is} = \frac{1}{N\,\pi_i}$$

and the sample size is fixed, $\hat{V}(t_5)$ reduces to the Yates and Grundy (56) form of the estimator of variance of the H.T. estimator.

F. Order in Stratification

It is well known that if a finite population is properly stratified, considerable gains in precision may be obtained with the stratified design. Neyman (41) and others have discussed the aspects of stratification in detail. Dalenius (7) and Dalenius and Hodges (8) determined the optimum stratum points in order to minimize the variance of the mean. They assumed that it was possible to stratify with respect to the characteristic under study. In the following, we shall assume a given random variable X with known bounded density function f(x). The random variable X is assumed to have a finite range with a variance denoted σ^2.

If we select a random sample of size n from a distribution with the density f(x), the variance of the sample mean, $\bar{X}_n$ is σ^2/n. We now show the order of decrease of the variance of $\bar{X}_n$ when we stratify the population into n strata.

Theorem VI.8:

Given a bounded random variable X with density function

$$f(x) \geq 0 \quad \text{when } x \varepsilon [a,b] \quad b > a$$
$$= 0 \quad \text{otherwise} .$$

Let n be given. Define the points a_i $i=1,\ldots,n$ by

$$\int_{a_{i-1}}^{a_i} f(x)dx = \frac{1}{n} .$$

Let stratum i be defined by the interval $[a_{i-1}, a_i]$. Then if we sample one per stratum, independently from stratum to stratum, $V(\bar{X}_n)$ is of order $1/n^2$.

Proof:

We have

$$E[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} \frac{xf(x)}{1/n} dx = \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} xf(x)dx = \int_a^b xf(x)dx$$

$$= E[X] \quad .$$

Also,

$$V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i)$$

where

$$V(X_i) = \int_{a_{i-1}}^{a_i} (x - E[X_i])^2 \frac{f(x)}{1/n} dx$$

$$\leq (a_i - a_{i-1})^2 \int_{a_{i-1}}^{a_i} \frac{f(x)dx}{1/n} = (a_i - a_{i-1})^2$$

we have that

$$V(\bar{X}_n) \leq \frac{1}{n^2} \sum_{i=1}^{n} (a_i - a_{i-1})^2 \leq \frac{1}{n^2} (b-a)^2 .$$

Hence $V(\bar{X}_n)$ is of order $1/n^2$. Q.E.D.

We note that the result of Theorem VI.8 remains valid if we choose to select more than one unit per stratum. With further restrictions on the density function we have

Theorem VI.9:

If in Theorem VI.8

$$f(x) \geq C > 0 \quad x \varepsilon [a,b]$$

$$= 0 \quad \text{elsewhere,}$$

then $V(\bar{X}_n)$ is of order $1/n^3$.

Proof:

As in Theorem VI.8 we have

$$V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} (x - E[X_i])^2 \frac{f(x)dx}{1/n} .$$

For any stratum, say stratum i, we have $(a_i - a_{i-1})C \leq \frac{1}{n}$ and hence $(a_i - a_{i-1}) \leq \frac{1}{nC}$ and

$$(a_i - a_{i-1})^2 \leq \frac{1}{n^2 C^2} \quad .$$

It follows that

$$V(\bar{X}_n) \leq \frac{1}{n^2} \sum_{i=1}^{n} \int_{a_{i-1}}^{a_i} \frac{(x - E[X_i])^2}{1/n} f(x)\,dx \leq \frac{1}{n^2} \sum_{i=1}^{n} \frac{1}{n^2 C^2}$$

$$\leq \frac{1}{n^3 C^2} \quad . \qquad \text{Q.E.D.}$$

If X is distributed as a uniform random variable and if we select a random sample of size n,

$$V(\bar{X}_n) = \frac{(b-a)^2}{12n} \quad .$$

However, if instead of selecting a random sample, we first stratify the population into n equal sized strata with respect to the uniform variable, it can easily be shown that

$$V(\bar{X}_n) = \frac{(b-a)^3}{3n^3} \quad .$$

Hence the variance of $\bar{X}_n$ utilizing stratification is $\frac{1}{n^2}$ times the variance of the simple mean with random sampling. We give an illustration of the usefulness of stratification in Theorem VI.10.

Theorem VI.10:

Given a finite population of N tuples (y_i, x_i) where $\bar{X}_N$ is known. Given also that

i) the density function of x satisfies the conditions of Theorem VI.8

ii) we stratify the finite population of N tuples into n strata of equal size by the method of Theorem VI.8 and

iii) we select one tuple per stratum from each of the n stratum with equal probability.

Then, omitting terms of order in probability $1/n^2$ we have that

$$\bar{V}(\bar{y}_n + \hat{\beta}(\bar{X}_N - \bar{x}_n)) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{L} \frac{(u_{(i)j} - \bar{U}_{(i)})^2}{N}$$

where

i) $$\hat{\beta} = \frac{\sum^{n} (x_i - \bar{x}_n) y_i}{\sum^{n} (x_i - \bar{x}_n)^2}$$

ii) $u_{(i)j} = y_{(i)j} - \bar{Y}_N - B(x_{(i)j} - \bar{X}_N)$ and

iii) $$\bar{U}_{(i)} = \frac{1}{L} \sum_{j=1}^{L} u_{(i)j}$$

The notation $y_{(i)j}$ denotes the value y of the j^{th} unit in the i^{th} stratum.

Proof:

Assume that $N = nL$. Define

$$B = \frac{\sum_{i=1}^{n} \sum_{j=1}^{L} (x_{(i)j}-\bar{X}_N)y_{(i)j}}{\sum_{i=1}^{n} \sum_{j=1}^{L} (x_{(i)j}-\bar{X}_N)^2}$$

and $u_{(i)j}$ as in ii) above. It follows that

$$\sum_{i=1}^{n} \sum_{j=1}^{L} u_{(i)j} = 0 = \sum_{i=1}^{n} \sum_{j=1}^{L} (u_{(i)j})(x_{(i)j}-\bar{X}_N)$$

It is easy to show that

$$\bar{y}_n + \hat{\beta}(\bar{X}_N - \bar{x}_n) - \bar{Y}_N = (\hat{\beta} - B)(\bar{X}_N - \bar{x}_n) + \bar{u}_n \quad . \qquad \text{(VI.3)}$$

Utilizing the sequence of populations model it can be shown that $\hat{\beta} - B = O_p(1/\sqrt{n})$. Hence since by Theorem VI.8,

$$\bar{X}_N - \bar{x}_n = O_p(\frac{1}{n})$$

we have from Equation VI.3 that

$$(\bar{y}_n + \hat{\beta}(\bar{X}_N - \bar{x}_n) - \bar{Y}_N)^2 = (\bar{u}_n + O_p(\frac{1}{n^{3/2}}))^2 = \bar{u}_n{}^2 + O_p(\frac{1}{n^2}) \quad .$$

Hence omitting terms of order in probability $1/n^2$

$$\bar{V}(\bar{y}_n + \hat{\beta}(\bar{X}_N - \bar{x}_n) - \bar{Y}_N) = \bar{E}(\bar{u}_n{}^2)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{L} \frac{(u_{(i)j}-\bar{U}_{(i)})^2}{N} \quad . \qquad \text{Q.E.D.}$$

VII. SUMMARY

Using the criterion of Anticipated Variance (A-V), sampling designs and estimators possessing desirable properties for certain models were derived.

We introduced the term Anticipated Variance to emphasize the subjective or personalistic nature of survey design. In practice, the determination of sampling design and estimator is often based on the survey designer's experience and intuition concerning a single finite population. As a result of previous surveys, the survey designer possesses prior knowledge concerning the population of interest. In the thesis, prior knowledge in the form of models specifying the anticipated means and covariance is investigated. The A-V criterion is broader than the Expected Variance criterion in that the former will apply when the a priori information does not arise as a result of a superpopulation but arises from a personalistic view of the finite population.

When the first and second order moments are assumed constant the sample mean together with simple random sampling without replacement were found to be best in that they minimized the A-V over a large class of estimators and sampling designs. Generalizing the moments we proved the following theorem.

Theorem:

Given that the anticipated mean of unit i is x_i, the variance is γ_i and that the variables $(y_1,\ldots,y_N)$ are independent with respect to some distribution. Furthermore, assume that $\bar{X}_N$ and the ratios $\sqrt{\gamma_i}/\sqrt{\gamma_j}$ $\forall i \neq j$ are known. Then, in the class of without replacement designs with the sample consisting of tuples (y,x) and expected sample size

$$n = \sum^{N} \pi_i$$

and in the class of unbiased estimators, the unequal probability difference estimator,

$$d_1 = \sum_{i\varepsilon s} \frac{y_i - x_i}{N\pi_i} + \bar{X}_N$$

and the sampling design with

$$\pi_i = \frac{n\sqrt{\gamma_i}}{\overset{N}{\Sigma}\sqrt{\gamma_i}}$$

jointly minimize the A-V. In connection with the model, classes of estimators and designs in the previous theorem we generalized the lower bound theorem of Godambe and Joshi (17). We also showed that the estimator d_1, is unbiased admissible, i.e., d_1 is admissible with respect to the class of unbiased estimators.

When the a priori mean is x_i for unit i, the variance is $\mathcal{V}_i$ and the covariance is ρ_{ij} for $i \neq j$, a programming problem was presented. The class of estimators was restricted to unbiased estimators of the form

$$\sum_{i=1}^{n} \beta_{is} y_i$$

where β_{is} is the coefficient of unit i whenever it is in the s^{th} sample and the class of sampling designs was restricted to without replacement designs of size n. For large populations such a method of solution would be prohibitive with respect to cost.

Three other models were considered. With respect to all three models, consistent estimators of the population mean were derived. When the a priori variance of unit i was $\mathcal{V}_i$ and the a priori covariance of units i and j was zero, the A-V of the consistent estimators omitting terms of order in probability larger than $\frac{1}{n}$ was minimized for π_i proportional $\sqrt{\mathcal{V}_i}$.

When the anticipated mean was of the form βx_i with β unknown, the anticipated variance was σ^2 and the covariances were zero we considered three different mixtures of two estimators with desirable properties. For each of the three mixture estimators we showed that the mixture estimator possessed a smaller anticipated variance than the usual ratio estimator

$$\frac{\bar{y}_n}{\bar{x}_n} \bar{X}_N \quad .$$

Consistency of the Horwitz-Thompson estimator was established for a sequence of finite populations. Several miscellaneous topics concerning sampling designs and estimators were considered. A class of variable sample size designs was shown to be inferior to the class of without replacement designs of size n admitting a nonnegative estimator of variance. This result assumed that the population characteristic of interest was positive.

Stratification in with replacement sampling was also considered. We showed that if it was possible to stratify the units into strata such that the values np_i sum to integers, the resulting design had a smaller variance than the usual with replacement unequal probability design of size n. Further topics included a discussion of the order in probability of $\bar{x}_n - \bar{X}_N$ associated with stratification where the number of strata is proportional to n. Also, a simple cost function of the form

$$\sum^{N} c_i \pi_i$$

where π_i is the inclusion probability of unit i, was considered. With respect to this cost function, the π_i that minimized the A-V of the unequal probability difference estimator was found to be

$$\pi_i = \frac{C_o\sqrt{\gamma_i/c_i}}{\sum^{N} \sqrt{\gamma_i c_i}} \quad .$$

VIII. BIBLIOGRAPHY

1. Aggarwal, O. P. Bayes and minimax procedures in sampling from finite and infinite populations. Institute of Mathematical Statistics Journal 30: 353-375. 1959.

2. Basu, D. On sampling with and without replacement. Sankhyā 20: 287-294. 1958.

3. Blackwell, D. and Girschick, M. A. Theory of games and statistical decisions. John Wiley and Sons, New York, New York. 1954.

4. Cochran, W. G. Relative accuracy of systematic and stratified random samples for a certain class of populations. Institute of Mathematical Statistics Journal 17: 164-177. 1946.

5. Cochran, W. G. Sampling techniques. Second edition. John Wiley and Sons, New York, New York. 1963.

6. Cramer, H. Mathematical methods of statistics. Princeton University Press, Princeton, New Jersey. 1963.

7. Dalenius, T. The problem of optimum stratification. Skandinavisk Aktuarietidskrift 3-4: 203-213. 1950.

8. Dalenius, T. and Hodges, J. L. Minimum variance stratification. American Statistical Association Journal 54: 89-100. 1959.

9. Durbin, J. Some results in sampling theory when the units are selected with unequal probabilities. Royal Statistical Society Journal 15: 262-269. 1953.

10. Ericson, W. A. Optimum stratified sampling using prior information. American Statistical Association Journal 60: 750-771. 1961.

11. Fisz, Marek. Probability theory and mathematical statistics. John Wiley and Sons, New York, New York. 1963.

12. Fellegi, I. P. Sampling with varying probabilities without replacement. American Statistical Association Journal 58: 183-201. 1963.

13. Fuller, W. A. Class notes for Statistics 638 course. Department of Statistics, Iowa State University, Ames, Iowa. 1969.

14. Fuller, W. A. A procedure for selecting nonreplacement unequal probability samples. Unpublished manuscript. Department of Statistics, Iowa State University, Ames, Iowa. 1970.

15. Fuller, W. A. Random stratum boundaries. Unpublished manuscript. Department of Statistics, Iowa State University, Ames, Iowa. ca. 1969.

16. Godambe, V. P. A unified theory of sampling from finite populations. Royal Statistical Society Journal Series B, 17: 269-278. 1955.

17. Godambe, V. P. and Joshi, V. M. Admissibility and Bayes estimation in sampling finite populations, I. Institute of Mathematical Statistics Journal 36: 1707-1722. 1965.

18. Goodman, L. A. and Hartley, H. O. The precision of unbiased ratio-type estimators. American Statistical Association Journal 53: 490-508. 1958.

19. Hadley, G. Nonlinear and dynamic programming. Addison-Wesley, Reading, Massachusetts. 1964.

20. Hájek, J. Optimum strategy and other problems in probability sampling. Časopis Pro Pestovani Matematiky 84: 387-423. 1959.

21. Hájek, J. Some contributions to the theory of probability sampling. Institut International de Statistique Bulletin 36: 127-134. 1958.

22. Hanurav, T. V. Hyper-admissibility and optimum estimator for sampling finite populations. Institute of Mathematical Statistics Journal 39: 621-642. 1968.

23. Hanurav, T. V. On Horwitz and Thompson estimator. Sankhyā Series A, 24: 429-436. 1962.

24. Hanurav, T. V. Optimum utilization of auxiliary information: πps sampling of two units from a stratum. Royal Statistical Society Journal, Series B, 29: 374-391. 1967.

25. Hanurav, T. V. Some aspects of unified sampling theory. Sankhyā Series A, 28: 175-204. 1966.

26. Hartley, H. O. and Rao, J. N. K. Sampling with unequal probabilities and without replacement. Institute of Mathematical Statistics Journal 33: 350-374. 1962.

27. Hartley, H. O. and Ross, A. Unbiased ratio estimators. Nature 174: 270-271. 1954.

28. Horwitz, D. G. and Thompson, D. J. A generalization of sampling without replacement from a finite universe. American Statistical Association Journal 47: 663-685. 1952.

29. Huntsberger, D. V. An extension of preliminary tests of significance permitting control of disturbances in statistical inferences. Unpublished Ph.D. thesis. Library, Iowa State University, Ames, Iowa. 1954.

30. Joshi, V. M. Admissibility and Bayes estimation in sampling finite populations, IV. Institute of Mathematical Statistics Journal 37: 629-638. 1966.

31. Koop, J. C. On the axioms of sample formation and their bearing on the construction of linear estimators in sampling theory for finite universes. Metrika 7: 81-114. 1963.

32. Lahiri, D. B. A method for sample selection providing unbiased ratio estimates. Institut International de Statistique Bulletin 33: 133-140. 1951.

33. Madow, W. G. On the limiting distributions of estimates based on samples from finite universes. Institute of Mathematical Statistics Journal 19: 535-545. 1945.

34. Mahalanobis, P. C. Some aspects of the design of surveys. Sankhyā 12: 1-7. 1952.

35. Mann, H. B. and Wald, A. On stochastic limit and order relationships. Institute of Mathematical Statistics Journal 14: 217-226. 1943.

36. Mickey, M. R. Some finite population unbiased ratio and regression estimators. American Statistical Association Journal 54: 594-612. 1959.

37. Midzuno, H. On the sampling system with probability proportionate to sum of sizes. Institute of Statistical Mathematics Journal 2: 99-108. 1951.

38. Mosteller, F. On pooling data. American Statistical Association Journal 43: 231-242. 1948.

39. Murthy, M. N. Ordered and unordered estimators in sampling without replacement. Sankhyā A, 18: 379-390. 1957.

40. Narain, R. D. On sampling without replacement with varying probabilities. Indian Society of Agricultural Statistics Journal 3: 169-174. 1951.

41. Neyman, J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Royal Statistical Society Journal 97: 558-606. 1934.

42. Nieto, J. Unbiased ratio estimators in stratified sampling. American Statistical Association Journal 56: 70-87. 1961.

43. Pathak, P. K. Sufficiency in sampling theory. Institute of Mathematical Statistics Journal 35: 795-809. 1964.

44. Pearson, K. Tables of the incomplete Γ-function. Cambridge University Press, London, England. 1951.

45. Pratt, J. W., Raiffa, H. and Schlaifer, R. The foundations of decision under uncertainty: an elementary exposition. American Statistical Association Journal 59: 353-375. 1964.

46. Raj, D. Sampling theory. McGraw-Hill, New York, New York. 1968.

47. Rao, J. N. K. On the precision of Mickey's unbiased ratio estimator. Biometrika 54: 321-324. 1967.

48. Rao, J. N. K., Hartley, H. O., and Cochran, W. G. On a simple procedure of unequal probability sampling without replacement. Royal Statistical Society Journal, Series B, 24: 482-491. 1962.

49. Rosenzweig, M. S. Ordered estimators for skewed populations. Unpublished Ph.D. thesis. Library, Iowa State University, Ames, Iowa. 1968.

50. Roy, J. and Chakravarti, I. M. Estimating the mean of a finite population. Institute of Mathematical Statistics Journal 31: 392-398. 1960.

51. Sampford, M. R. On sampling without replacement with unequal probabilities of selection. Biometrika 54: 499-513. 1967.

52. Sukhatme, B. V. A comparison of two sampling procedures with an application to successive sampling. Unpublished manuscript. Department of Statistics, Iowa State University, Ames, Iowa. 1970.

53. Sukhatme, B. V. Some ratio-type estimators in two-phase sampling. American Statistical Association Journal 57: 628-632. 1962.

54. Vijayan, K. An exact πps sampling scheme - generalization of a method of Hanurav. Royal Statistical Society Journal, Series B, 30: 556-566. 1968.

55. Williams, W. H. Generating unbiased ratio and regression estimators. Biometrics 17: 267-274. 1961.

56. Yates, F. and Grundy, P. M. Selection without replacement from within strata with probability proportional to size. Royal Statistical Society Journal, Series B, 15: 253-261. 1953.

IX. ACKNOWLEDGMENT

The author gratefully acknowledges Professor Wayne A. Fuller for his guidance and encouragement during the preparation of this dissertation. He also acknowledges his wife, Leslie, for her patience and support.

X. APPENDIX 1

We present a proof for Theorem A.1.

Theorem A.1:

Given the model

$$\mathcal{E}(y_i \mid i) = \mu$$

$$\mathcal{C}ov(y_i, y_j \mid i, j) = \sigma^2 \qquad i = j$$
$$= \rho\sigma^2 \qquad i \neq j$$

where μ, σ^2, ρ are unknown. Then, any sampling design such that the inclusion probability $\pi_i = n/N \ \forall_i$ will guarantee that $\bar{y}_n$ is an unbiased estimator of $\bar{Y}_N$. Furthermore,

$$\text{A-V}(\bar{y}_n) = \sigma^2(\frac{1}{n} - \frac{1}{N}) + \rho\sigma^2(\frac{n-1}{n} - \frac{N-1}{N}) \quad .$$

Proof:

Since $\bar{y}_n$ is required to be unbiased, we have

$$E[\bar{y}_n] = \sum_{s\varepsilon S} \bar{y}_n p(s) = \sum_{i=1}^{N} \frac{y_i}{n} \sum_{s\ni i} p(s) = \bar{Y}_N \ \forall \text{ values } (y_1, \ldots, y_N).$$

This implies that

$$\sum_{s\ni i} p(s) = \frac{n}{N} \quad \forall_i$$

and hence

$$\pi_i = \frac{n}{N} \quad \forall_i \quad .$$

It follows that

$$\text{A-V}(\bar{y}_n) = E\mathscr{V}(\bar{y}_n - \bar{Y}_N | n) + V\xi(\bar{y}_n - \bar{Y}_N | s)$$

$$= E[\sigma^2(\frac{1}{n} - \frac{1}{N}) + \rho\sigma^2(\frac{n-1}{n} - \frac{N-1}{N})]$$

$$= \sigma^2(\frac{1}{n} - \frac{1}{N}) + \rho\sigma^2(\frac{n-1}{n} - \frac{N-1}{N}) \quad .$$

XI. APPENDIX 2

We present a restatement and proof of Theorem A.2.

Theorem A.2:

Given the specifications of Model I. Given that the sample design is simple random sampling without replacement of size n (SRS w/o replacement of size n). Then, the sample mean $\bar{y}_n$, is such that no unbiased estimator in the class T_7 has smaller A-V. The class

$$T_7 = \{t_7 \colon t_7 = \sum_{i=1}^{n} c_{s_{i_t}} y_i \quad \text{where} \quad c_{s_{i_t}}$$

is the coefficient of element U_i whenever it appears at the t^{th} draw in the s^{th} sample$\}$.

Proof:

We first note that the result of Murthy (39) and others enables us to consider estimators in T_5 instead of the larger class T_7. Since T_7 estimators take into account the order in which the units appear in the sample, there are $n!\binom{N}{n}$ possible samples. If we average t_7 over all possible $n!$ samples containing the same units U_i the resulting averaged estimator will not depend on the order in which the units appear and will be in T_5. Given that one of the $n!\binom{N}{n}$ samples, s', is drawn, and given t_7, we can construct the estimator

$$t_5 = \frac{\sum\limits_{s'\varepsilon S_0} t_7(s')p(s')}{\sum\limits_{s'\varepsilon S_0} p(s')}$$

where S_0 denotes the set of all samples s containing the same units U_i in s'.

Furthermore,

$$E[t_5] = \sum_{S_0} t_5 p(S_0) = \sum_{s'\varepsilon S} t_7(s')p(s') = E[t_7] \quad .$$

Now,

$$V(t_7) = EV(t_7|S_0) + VE[t_7|S_0]$$

where $V(t_7|S_0)$ denotes the variance of t_7 over all samples s' in the set S_0 and by construction $E[t_7|S_0] = t_5$. Hence,

$$V(t_7) \geq V(t_5) \quad .$$

Having shown that we may limit our consideration to the class T_5, let

$$t_5 = \sum_{i=1}^{n} \beta_{is} y_i \quad .$$

We now calculate and minimize the A-V(t_5). The

$$\mathscr{E}(t_5|s) = \sum_{i=1}^{n} \beta_{is}\mu$$

$$\mathscr{V}(t_5|s) = \sum_{i=1}^{n} \beta_{is}^2\sigma^2 + \sum_{i\neq j}^{n} \beta_{is}\beta_{js}\rho\sigma^2$$

and

$$\mathcal{C}ov(t_5, \bar{Y}_N | s) = \mathcal{E}[(\sum^{n} \beta_{is} y_i)(\frac{1}{N} \sum_{i=1}^{N} (y_i - \mu))]$$

$$= \frac{1}{N} \sum_{i=1}^{n} \beta_{is} \sigma^2 + \frac{1}{N} \sum_{i \not\in s} \beta_{is} \rho \sigma^2$$

$$= \frac{1}{N} \sum_{i=1}^{n} \beta_{is} \sigma^2 + \frac{N-1}{N} \sum_{i=1}^{n} \beta_{is} \rho \sigma^2 .$$

$$A\text{-}V(t_5) = V(\sum_{i=1}^{n} \beta_{is} \mu - \mu) + E[\sum_{i=1}^{n} \beta_{is}{}^2 \sigma^2 + \sum_{i \neq j}^{n} \beta_{is} \beta_{js} \rho \sigma^2$$

$$- \frac{2}{N}(\sum \beta_{is} \sigma^2 + N-1 \sum_{i \in s} \beta_{is} \rho \sigma^2) + \frac{\sigma^2}{N} + \frac{N-1}{N} \rho \sigma^2] \quad .$$

$$\geq E[\sum_{i=1}^{n} \beta_{is}{}^2 \sigma^2 + \sum_{i \neq j} \beta_{is} \beta_{js} \rho \sigma^2 - \frac{2}{N}(\sum \beta_{is} \sigma^2$$

$$+ (N-1) \sum_{i=1}^{n} \beta_{is} \rho \sigma^2)] + \frac{\sigma^2}{N} + (\frac{N-1}{N}) \rho \sigma^2$$

$$\geq E[\sum_{i=1}^{n} \beta_{is}{}^2 \sigma^2 + \sum_{i \neq j}^{n} \beta_{is} \beta_{js} \rho \sigma^2] - \frac{2}{N} \sigma^2 - \frac{2(N-1)}{N} \rho \sigma^2$$

$$+ \frac{\sigma^2}{N} + (\frac{N-1}{N}) \rho \sigma^2$$

$$\geq E[\sum_{i=1}^{n} \beta_{is}{}^2 \sigma^2 + \sum_{i \neq j}^{n} \beta_{is} \beta_{js} \rho \sigma^2] - \frac{\sigma^2}{N} - \frac{N-1}{N} \rho \sigma^2$$

$$\geq \sum_s \left(\sum_{i=1}^{n} \beta_{is}^2\sigma^2\right)p(s) + \sum_s\left(\sum_{i\neq j}^{n} \beta_{is}\beta_{js}\rho\sigma^2\right)p(s)$$

$$-\frac{\sigma^2}{N} - \left(\frac{N-1}{N}\right)\rho\sigma^2.$$

Let

$$H = \sum_s\left(\sum \beta_{is}^2\sigma^2\right)p(s) + \sum_s\left(\sum \beta_{is}\beta_{js}\rho\sigma^2\right)p(s) - \frac{\sigma^2}{N}$$

$$-\left(\frac{N-1}{N}\right)\rho\sigma^2 - \sum_{i=1}^{N} \lambda_i\left(\sum_{s\ni i} \beta_{is}p(s) - \frac{1}{N}\right)$$

where λ_i, i=1,...,N are Lagrange multipliers. Then,

$$\frac{\partial H}{\partial\beta_{is}} = 2\beta_{is}\sigma^2 p(s) + p(s)\rho\sigma^2\, 2\sum_{\substack{j=1\\ \neq i}}^{n} \beta_{is} - \lambda_i p(s) = 0 \qquad \text{(A.2.1)}$$

Adding and subtracting $2p(s)\rho\sigma^2\beta_{is}$, we have

$$\frac{\partial H}{\partial\beta_{is}} = 2\beta_{is}\sigma^2 p(s) + p(s)\rho\sigma^2 2\sum_{i=1}^{n} \beta_{is} - 2p(s)\rho\sigma^2\beta_{is} - \lambda_i p(s) = 0\ .$$

(A.2.2)

Now summing A.2.2 over all s containing i, we have

$$\frac{2\sigma^2}{N} + 2\rho\sigma^2 \sum_{s\ni i}\left(\sum_{i=1}^{n} \beta_{is}\right)p(s) - 2\rho\frac{\sigma^2}{N} - \lambda_i\frac{n}{N} = 0$$

and hence

$$\hat{\lambda}_i = -\frac{1}{n}\left[2\rho\sigma^2 - 2\sigma^2 - \rho\sigma^2 N2\sum_{s\ni i}\left(\sum_{i=1}^{n} \beta_{is}\right)p(s)\right]\ .$$

Substituting into Equation A.2.1 we have

$$2\beta_{is}\sigma^2 + \rho\sigma^2\ 2\sum_{\substack{j=1\\ \neq i}}^{n} \beta_{js} + \frac{1}{n}[2\rho\sigma^2 - 2\sigma^2 - \rho\sigma^2 N2 \sum_{s \ni i} (\sum^{n} \beta_{is})p(s)] = 0 \ . \tag{A.2.3}$$

The values $\beta_{is} = 1/n\ \forall_i$, s are solutions to Equation A.2.3. The expression

$$\sum_{i=1}^{n} (\beta_{is})^2\sigma^2 + \rho\sigma^2 \sum_{i \neq j}^{n} \beta_{is}\beta_{js} = \mathcal{V}(t_5|s) \geq 0 \ .$$

Also $\mathcal{V}(t_5|s)$ is convex in β_{is} for any fixed s. This is because $\mathcal{V}(t_5|s)$ can be written as a positive semidefinite quadratic form in the β_{is}. Hence

$$\sum_{s} \mathcal{V}(t_5|s)p(s) = \text{A-V}(t_5)$$

is convex in β_{is} for $i = 1,\ldots,N$ and for all s. By the Kuhn-Tucker Sufficiency Theorem given on page 192, Hadley (19), the A-V(t_5) with $\beta_{is} = \frac{1}{n}$ is the global minimum over the closed convex set given by

$$\beta_{is} \in R^1 \quad \forall_{i,s}$$

and

$$\sum_{s \ni i} \beta_{is}p(s) = \frac{1}{N} \quad \forall_i \ .$$

XII. APPENDIX 3

We present a restatement of Theorem A.3 and base our proof on a result of Hájek (20).

Theorem A.3:

In Model I, if $\rho \neq 0$, $\bar{y}_n$ and SRS w/o replacement of size n are jointly best for unbiased estimators in the class T_7 and fixed sample size designs.

Proof:

From Appendix 2 we restrict outselves to T_5. Hájek (20) showed that for Model III, unequal probability systematic sampling, such that π_i is proportional to x_i and the Horwitz-Thompson estimator are jointly best. Model III is of the form

$$\mathscr{E}(y_i|x_i) = x_i > 0$$

$$\begin{aligned} \mathscr{C}ov(y_i, y_j|x_i x_j) &= cx_i^2 & i = j \\ &= \sigma^2 x_i x_j R(|i-j|) & i \neq j \end{aligned}$$

where the x_i are known prior to sampling and R is a convex function of $|i-j|$.

We note that Model I is a special case of Model III where x_i = constant $\forall_i$ and R is a constant function. Hence applying Hajek's result, the sample mean $\bar{y}_n$ together with equal probability systematic sampling jointly minimize the

A-V(t_5). Also, given Model I, the A-V($\bar{y}_n$) under SRS w/o replacement yields the same A-V($\bar{y}_n$) as the equal probability systematic design. In practice, SRS w/o replacement is to be preferred since it allows for the construction of an unbiased estimator of variance.

XIII. APPENDIX 4

Under the model

$$\mathcal{E}(y_i | x_i) = x_i$$

$$\mathcal{C}ov(y_i, y_j | x_i, x_j) = 0 \qquad i \neq j$$

$$= \mathcal{V}_i \qquad i = j$$

Godambe and Joshi (17) gave a lower bound for the A-V of estimators that are unbiased for $\bar{Y}_N$. They assumed that the sample consisted of the characteristic y_i only and that y_i, $i=1,\ldots,N$ were independent with respect to the distribution generating the y_i.

In Theorem A.4 below, we generalize their result by assuming sampling of tuples (y_i, x_i) where x_i is a concomittant variable that is observed when y_i is observed and

$$\bar{X}_N = \sum_{i=1}^{N} \frac{x_i}{N}$$

is known prior to sampling. The format of the proof is essentially that of Godambe and Joshi (17).

Theorem A.4:

Given that a sample consists of tuples (y_i, x_i), given $\bar{X}_N$ is known and given the population characteristics y_i, $i=1,\ldots,N$ are independent with respect to some underlying distribution where

$$\mathscr{E}(y_i|x_i) = x_i$$

$$\mathscr{C}ov(y_i,y_j|x_i,x_j) = 0 \qquad i \neq j$$
$$= \mathscr{V}_i \qquad i = j \ ,$$

and given $e(s,(y,x))$, any unbiased estimator of $\bar{Y}_N$ that is a function of the sample s and the observed values $(y_i,x_i)\varepsilon s$. Then,

$$\text{A-V}(e(s,(y,x)) \geq \sum_{i=1}^{N} \frac{\mathscr{V}_i}{N^2\pi_i} - \sum_{i=1}^{N} \frac{\mathscr{V}_i}{N^2}$$

for any sampling design such that the inclusion probability, π_i, of every unit in the population is positive.

Proof:

Let $e(s,(y,x)) = d_1 + h(s,(y,x))$ be any unbiased estimator that is a function of (y_i,x_i) from the sample s, $\bar{X}_N$ and of the units U_i that compose the sample where

$$d_1 = \sum_{i\varepsilon s} \frac{y_i}{N\pi_i} + \bar{X}_N - \sum_{i\varepsilon s} \frac{x_i}{N\pi_i} = \sum_{i\varepsilon s} \frac{y_i - x_i}{N\pi_i} + \bar{X}_N \ .$$

For ease of notation, we write $e = d_1 + h$. Given a sample s,

$$\mathscr{C}ov(d_1,h|s) = \mathscr{E}[(d_1 - \mathscr{E}(d_1|s))(h - \mathscr{E}(h|s))]$$

$$= \mathscr{E}[(\sum_{i\varepsilon s} \frac{y_i - x_i}{N\pi_i})h|s]$$

since $\mathscr{E}(d_1|s) = \bar{X}_N$.

The sampling variance of e is

$$V(e) = V(d_1) + V(h) + 2\ \mathrm{Cov}(d_1,h)$$

and the

$$\text{A-V}(e) = \text{A-V}(d_1) + \text{A-V}(h) + 2\,\mathcal{E}\,\mathrm{Cov}(d_1,h)$$

$$\geq \text{A-V}(d_1) + 2\,\mathcal{E}\,\mathrm{Cov}(d_1,h) \quad . \qquad \text{(A.4.1)}$$

We now show that $\mathcal{E}\,\mathrm{Cov}(d_1,h) = 0$. As both d_1 and e are unbiased for $\bar{Y}_N$, $E(h) = 0$ and we have

$$\mathrm{Cov}(d_1,h) = E[d_1\ h] - E[d_1]E[h]$$

$$= E[d_1 h]$$

$$= \sum_s p(s)\left(\sum_{i\varepsilon s} \frac{y_i - x_i}{N\pi_i} + \bar{X}_N\right)h(s)$$

$$= \sum_s p(s)\left(\sum_{i\varepsilon s} \frac{y_i - x_i}{N\pi_i}\right)h(s)$$

and hence,

$$\mathcal{E}\,\mathrm{Cov}(d_1,h) = \mathcal{E}\sum_s p(s)\left(\sum_{i\varepsilon s} \frac{y_i - x_i}{N\pi_i}\right)h(s)$$

$$= \mathcal{E}\sum_{i=1}^{N}\left(\frac{y_i - x_i}{N\pi_i}\right)\sum_{s\ni i} p(s)h(s) \quad .$$

By the unbiasedness restriction

$$E[h(s)] = \sum_s h(s)p(s) = 0$$

$$\sum_{s \ni i} h(s)p(s) = - \sum_{s \not\ni i} h(s)p(s) \quad .$$

we have

$$\mathcal{E}\,\mathrm{Cov}(d_1,h) = - \mathcal{E} \sum_{i=1}^{N} \left(\frac{y_i - x_i}{N\pi_i}\right) \sum_{s \not\ni i} h(s)p(s) = 0 \quad .$$

The last equality holding by virtue of the independence of $(y_1,\ldots,y_N)$ with respect to the underlying distribution. From Equation A.4.1 we now have

$$\text{A-V}(e) \geq \text{A-V}(d_1) = \sum_{i=1}^{N} \frac{\mathcal{V}_i}{N^2\pi_i} - \sum_{i=1}^{N} \frac{\gamma_i}{N^2} \quad . \qquad \text{(A.4.2)}$$

We note that if $\mathcal{E}(y_i|x_i) = f(x_i)$ the lower bound in Equation A.4.2 remains the same. If the form of f were known prior to sampling the estimator d_1^* given by

$$d_1^* = \sum_{i \varepsilon s} \frac{y_i}{N\pi_i} + \left(\sum_{i=1}^{N} \frac{f(x_i)}{N} - \sum_{i \varepsilon s} \frac{f(x_i)}{N\pi_i}\right)$$

attains the lower bound.

XIV. APPENDIX 5

Godambe and Joshi (17) show that the Horwitz-Thompson estimator is admissible for any design that has positive inclusion probabilities for every unit in the population. Godambe and Joshi define e to be unbiased admissible[1] for $\bar{Y}_N$, for a given sampling design, if

i) e is unbiased for $\bar{Y}_N$ and

ii) there does not exist another unbiased estimator e^* such that $V(e^*) \leq V(e)$ no matter what the values $(y_1,\ldots,y_N)$ and strict inequality holding for at least one set of values $(y_1,\ldots y_N)$.

We extend the result of Godambe and Joshi in Theorem A.5.

Theorem A.5.

Given that a sample s consists of tuples (y_i,x_i), $i\varepsilon s$, $\bar{X}_N$ is known and $(x_1,\ldots x_N)$ is a fixed point in R^N, the estimator $d_1 - \bar{X}_N$ (d_1 as defined in Appendix 4) is unbiased admissible for $\bar{Y}_N - \bar{X}_N$, and hence d_1 is unbiased admissible for $\bar{Y}_N$.

Proof:

The proof follows closely that of Godambe and Joshi (17). First, assume that $d_1 - \bar{X}_N$ is not admissible for

[1]Godambe and Joshi used the term admissible. We have adopted the term unbiased admissible to avoid confusion with admissibility as it is commonly used in decision theory.

$\bar{Y}_N - \bar{X}_N$. Then there exists an unbiased estimator

$$e(s_1(y,x)) = d_1 - \bar{X}_N + h(s,(y,x))$$

that has at least as small variance as $d_1 - \bar{X}_N$ for any values of $(y_1,\ldots,y_N)$ and has smaller variance for at least one value of $(y_1,\ldots,y_N)$. In other words, the estimator e must be such that

$$\sum_{s\varepsilon S} h(s)p(s) = 0 \qquad \text{(A.5.1)}$$

and

$$\sum_{s\varepsilon S} e^2 p(s) \leq \sum_{s} (d_1 - \bar{X}_N)^2 p(s) \text{ for all } (y_1,\ldots,y_N)\varepsilon R^N \qquad \text{(A.5.2)}$$

or

$$\sum_{s\varepsilon S} h^2 p(s) \leq -2 \sum_{s\varepsilon S} h(d_1 - \bar{X}_N)p(s) \text{ for all } (y_1,\ldots,y_N)\varepsilon R^N$$

with strict inequality for at least one $(y_1,\ldots,y_N)\varepsilon R^N$. As in Appendix 4, we have abbreviated $e(s,(y,x))$ and $h(s,(y,x))$ to simplify expressions. The proof of Theorem A.5 consists of showing that $h \equiv 0$ for all samples s with $p(s) > 0$.

Let Y^k denote a subset of R^N and be such that if $y-x = (y_1-x_1,\ldots,y_N-x_N)$ is a member of Y^k, then just k values of y-x are nonzero.

Lemma:

If $h(s)p(s) = 0 \;\forall\; (s(y,x))$ such that $s\varepsilon S$ and $(y-x)\varepsilon Y^k$, then $h(s)p(s) = 0 \;\forall\; (s(y,x))$ such that $s\varepsilon S$ and $(y-x)\varepsilon Y^{k+1}$.

Proof:

Let $(y' - x)\varepsilon Y^{k+1}$. From Equation A.5.1 we have

$$\Sigma h(s)p(s) = 0 = \sum_{i=0}^{k+1} \sum_{s\varepsilon S_i} h(s,(y',x))p(s) \tag{A.5.3}$$

and from Equation A.5.2 we have

$$\sum_{i=0}^{k+1} \sum_{s\varepsilon S_i} h^2(s,(y',x))p(s) \leq -2 \sum_{i=0}^{k+1} \sum_{s\varepsilon S_i} h(s,(y',x))(d_1-\bar{X}_N)p(s) \tag{A.5.4}$$

where S_i is the collection of all samples s containing exactly i units with the property $y_1' - x_i \neq 0$. We note that summation on i is from 0 to k+1 only since $(y'-x)\varepsilon Y^{k+1}$ and there can be at most k+1 units in the sample such that $y_i' - x_i \neq 0$. Briefly then,

$$\bigcup_{i=0}^{k+1} S_i = S \quad .$$

We are given that $h(s)p(s) = 0$ for all $(s(y,x))$ such that $s\varepsilon S$, and $(y-x)\varepsilon Y^k$. This implies that $h(s,(y',x))p(s)=0$ for all S_i, $i=0,\ldots,k$. Thus from Equations A.5.3 and A.5.4 we have

$$\sum_{s \varepsilon S_{k+1}} h(s)p(s) = 0 \tag{A.5.5}$$

and

$$\sum_{s \varepsilon S_{k+1}} h^2(s,(y',x))p(s) \le -2 \sum_{s \varepsilon S_{k+1}} h(s,(y',x))(d_1 - \bar{X}_N)p(s) \ . \tag{A.5.6}$$

In Equation A.5.6,

$$d_1 - \bar{X}_N = \sum_{i \varepsilon s} \frac{y_i - x_i}{N\pi_i}$$

is constant for all samples $s \varepsilon S_{k+1}$ and since $(y'-x) \varepsilon Y^{k+1}$, for all samples $s \varepsilon S_{k+1}$,

$$d_1 - \bar{X}_N = \sum^{N} \frac{y_i' - x_i}{N\pi_i} \quad .$$

Hence by Equation A.5.5,

$$\sum_{s \varepsilon S_{k+1}} h^2(s,(y',x))p(s) \le -2[\sum_{i=1}^{N} \frac{y_i - x_i}{N\pi_i}] \sum_{s \varepsilon S_{k+1}} h(s,(y',x))p(s) = 0 \ .$$

Thus $h(s,(y',x))p(s) = 0$ for all $s \varepsilon S_{k+1}$. Since

$$S = \bigcup_{i=1}^{k+1} S_i \quad ,$$

$h(s,(y',x))p(s) = 0$ for all $s \varepsilon S$ and the lemma is proved.

For $(y-x) \varepsilon Y^{o}$, and all $s \varepsilon S$,

$$h(s,(y,x))p(s) = 0 \quad . \tag{A.5.7}$$

The equality follows because of the assumption that $d_1 - \bar{X}_N$ is inadmissible for $\bar{Y}_N - \bar{X}_N$. When $(y-x)\varepsilon Y^o$, $e(s,(y,x) = h(s,(y,x))$ and admissibility of e implies that $V(e) \leq V(d_1 - \bar{X}_N) = 0$. Thus when $(y-x)\varepsilon Y^o$, $e(s,(y,x)) = \text{constant}$ for samples s. Since $E[e] = \bar{Y}_N - \bar{X}_N$, this constant must be zero. Hence, using the lemma above with $k = 0$,

$$h(s,(y,x))p(s) = 0 \text{ for all } s\varepsilon S \text{ and } (y-x)\varepsilon R^N \quad .$$

Since x is a fixed point,

$$h(s,(y,x))p(s) = 0 \text{ for all } s\varepsilon S \text{ and } y\varepsilon R^N \quad .$$

Hence, we have a contradiction. We assumed that $V(e) < V(d_1 - \bar{X}_N)$ for at least one $(y_1,\ldots,y_N)\varepsilon R^N$ and have now shown this cannot occur.

Now, since $d_1 - \bar{X}_N$ is unbiased admissible for $\bar{Y}_N - \bar{X}_N$, it follows that d_1 is unbiased admissible for $\bar{Y}_N$.

XV. APPENDIX 6

Theorem E.2:

Consider the model

$$\mathcal{E}(y_i|x_i) = x_i + c$$

$$\mathcal{C}ov(y_i,y_j|x_i,x_j) = 0 \qquad i \neq j$$
$$= \mathcal{V}_i \qquad i = j$$

where $\bar{X}_N$ is known and c is unknown. Let L denote the class of linear estimators of the form

$$\ell = \sum_{i=1}^{n} \beta_{is} y_i + f(s)$$

where β_{is} is the coefficient of unit U_i whenever it is in the s^{th} sample and f(s) is a function of the sample s but not of the $y_i \varepsilon s$. Given that the sample s consists of n distinct units. Then of all estimators that are conditionally unbiased under the model, the estimator

$$\bar{X}_N + \sum_{i=1}^{n} \frac{y_i - x_i}{\mathcal{V}_i} / \sum_{i=1}^{n} \frac{1}{\mathcal{V}_i}$$

has the smallest variance under the model.

Proof:

The condition of unbiasedness under the model implies that

$$\mathcal{E}(\ell|s) = \mathcal{E}(f(s) + \sum_{i=1}^{n} \beta_{is}y_i|s) = f(s) + \sum_{i=1}^{n}\beta_{is}x_i + c\sum_{i=1}^{n}\beta_{is}$$

$$= \bar{X}_N + c \quad \forall\ s\epsilon S \quad .$$

We then have that

$$\sum_{i=1}^{n} \beta_{is} = 1 \quad \forall\ s$$

and

$$\sum_{i=1}^{n} \beta_{is}x_i + f(s) = \bar{X}_N \quad .$$

We now write ℓ in the form

$$\ell = \sum_{i=1}^{n} \beta_{is}(y_i-x_i) + \bar{X}_N \quad .$$

The

$$\mathcal{V}(\sum_{i=1}^{n} \beta_{is}(y_i-x_i) + \bar{X}_N|s) = \sum_{i=1}^{n} \beta_{is}^{\ 2}\mathcal{V}_i$$

is a convex function in β_{is}. Let

$$\phi = \sum_{i=1}^{n} \beta_{is}^{\ 2}\mathcal{V}_i + 2\lambda(\sum_{i=1}^{n} \beta_{is} - 1)$$

where λ is the Lagrangian multiplier. Taking the partial derivative with respect to β_{is},

$$\frac{\partial \emptyset}{\partial \beta_{is}} = 2 \beta_{is} \gamma_i + 2\lambda = 0 \qquad \text{(A.6.1)}$$

Hence,

$$2 \sum_{i=1}^{n} \beta_{is}^{2} \gamma_i + 2\lambda = 0$$

and

$$\hat{\lambda} = - \sum_{i=1}^{n} \beta_{is}^{2} \gamma_i \quad .$$

Substituting $\hat{\lambda}$ into

$$\frac{\partial \emptyset}{\partial \beta_{is}} = 0 \quad ,$$

we have

$$2\beta_{is} \gamma_i - 2 \sum_{i=1}^{n} \beta_{is}^{2} \gamma_i = 0 \quad . \qquad \text{(A.6.2)}$$

It is easy to see that

$$\hat{\beta}_{is} = \frac{1}{\gamma_i} / \sum_{i=1}^{n} \frac{1}{\gamma_i}$$

satisfies the Equation A.6.2. By the Kuhn-Tucker sufficiency theorem cited in Appendix 2, $\hat{\beta}_{is}$ is a global minimum over the set given by $\beta_{is} \varepsilon R^1 \ \forall_i$ and

$$\sum_{i-1}^{n} \beta_{is} = 1 \quad .$$

XVI. APPENDIX 7

We consider the modification of the mixture estimator $\hat{\theta}_2$ when σ^2 is not known mentioned in Chapter IV. We also consider a modification in $\hat{\theta}_2$ resulting in a change in the limiting variances.

The estimator $\hat{\sigma}^2$ of σ^2 is unbiased under the model. This is so, since,

$$\mathcal{E}(\hat{\sigma}^2|s) = \mathcal{E}\sum^{n}(\beta x_i + e_i - \beta x_i \frac{(\Sigma x_i e_i)x_i}{\Sigma x_i^2})^2/n-1$$

$$= \sum^{n}\{\sigma^2 - \frac{2x_i^2\sigma^2}{\Sigma x_i^2} + \frac{x_i^2\sigma^2}{\Sigma x_i^2}\}/n-1$$

$$= \sigma^2 \quad .$$

Furthermore, replacing σ^2 by $\hat{\sigma}^2$ in $\hat{\alpha}$, we have

$$\hat{\alpha} = 1/[1 + \frac{\sqrt{n}\ (\hat{B}-\hat{R})^2\ \bar{x}_n^{\,2}(\sum^{n} \frac{x_i^2}{n})^2}{\{\Sigma\frac{x_i^2}{n}\Sigma\frac{y_i^2}{n-1} - \Sigma\frac{x_iy_i}{n}\Sigma\frac{x_iy_i}{n}\}\ \Sigma\frac{(x_i-\bar{x}_n)^2}{n}}] \quad .$$

Each additional term can be represented by a constant plus a term of $O_p(\frac{1}{\sqrt{n}})$. Hence substituting $\hat{\sigma}^2$ for σ^2 does not alter the consistency of $\hat{\theta}_2$. The variance of $\hat{\theta}_2$ is altered in that the value of C is now

$$C = \frac{(B_N - R_N)^2 K}{\Sigma \frac{x_i^2}{N} \Sigma \frac{y_i^2}{N} - (\Sigma \frac{x_i y_i}{N})^2} .$$

We now consider another modification in $\hat{\alpha}$.

Using $\hat{\alpha}'$ from Equation IV.30 in

$\hat{\theta}_2 = \hat{\alpha}\hat{R}_2 + (1-\hat{\alpha})\hat{R}_1$

and denoting it $\hat{\theta}_2'$, we have

$$\hat{\theta}_2' - \bar{Y}_N = \hat{\alpha}'\hat{R}_2 + (1-\hat{\alpha}')\hat{R}_1 - \bar{Y}_N$$

$$= \bar{X}_N(B-R)\hat{\alpha}' + \bar{X}_N[\hat{\alpha}' \frac{\sum^n x_i u_i}{\Sigma x_i^2} + (1-\hat{\alpha}') \frac{\sum^n s_i}{\sum^n x_i}] .$$

When B = R,

$$(\hat{\theta}_2' - \bar{Y}_N)^2 = \bar{X}_N^2 [\hat{\alpha}'^2 (\frac{\Sigma x_i u_i}{\Sigma x_i^2})^2 + (1-\hat{\alpha}')^2 (\frac{\sum^n s_i}{\sum^n x_i})^2$$

$$+ 2\hat{\alpha}'(1-\hat{\alpha}') \frac{\sum^n x_i u_i}{\sum^n x_i^2} \frac{\sum^n s_i}{\sum^n x_i}] .$$

Now,

$$\hat{\alpha}' = \frac{1}{1 + n^{\frac{1}{2}+\varepsilon}\,\frac{(\hat{B}-\hat{R})^2}{\sigma^2}\left[\frac{\bar{X}_N{}^2 \sum^N \frac{x_i{}^2}{N}}{\sum^N \frac{(x_i-\bar{X}_N)^2}{N}} + O_p(\frac{1}{\sqrt{n}})\right]}$$

and when $B = R$, $(\hat{B}-\hat{R})^2 = O_p(1/n)$ so that we may write

$$\hat{a}' = \frac{1}{1 + O_p(\frac{1}{n^{\frac{1}{2}-\varepsilon}})\left[\frac{\bar{X}_N{}^2 \sum\frac{x_i{}^2}{N}}{\frac{\sum(x_i-\bar{X}_N)^2}{N}} + O_p(\frac{1}{\sqrt{n}})\right]} = \frac{1}{1 + O_p(\frac{1}{n^{\frac{1}{2}-\varepsilon}})}.$$

We also have

i) $\hat{\alpha}' = 1 + O_p(\frac{1}{n^{\frac{1}{2}-\varepsilon}})$

ii) $\hat{\alpha}'^2 = 1 + O_p(\frac{1}{n^{\frac{1}{2}-\varepsilon}})$

iii) $(1-\hat{\alpha}')^2 = O_p(\frac{1}{n^{1-2\varepsilon}})$

Hence, omitting expectation of terms of order in probability greater than $1/n$, we have

$$\bar{E}[\hat{\theta}_2' - \bar{Y}_N)^2] = \bar{X}_N{}^2\ \bar{E}[(\frac{\sum x_i u_i}{\sum x_i{}^2})^2] \doteq \frac{\bar{X}_N{}^2(1-\frac{n}{N})N}{(\sum^N \frac{x_i{}^2}{N})^2}\sum\frac{[(y_i-Bx_i)x_i]^2}{(N-1)n}.$$

When $B \neq R$,

$$\hat{\alpha}' = \frac{1}{1 + n^{\frac{1}{2}+\varepsilon}\left((B-R)^2 + O_p\left(\frac{1}{\sqrt{n}}\right)\right)\left(\frac{\bar{X}_N^{\,2}\sum^{N}\frac{x_i^{\,2}}{N}}{\sum\frac{(x_i-\bar{X}_N)^2}{N}} + O_p\left(\frac{1}{\sqrt{n}}\right)\right)} = O_p\left(\frac{1}{n^{\frac{1}{2}+\varepsilon}}\right)$$

and we have,

i) $1 - \hat{\alpha}' = 1 - O_p\left(\frac{1}{n^{\frac{1}{2}+\varepsilon}}\right)$

ii) $(1 - \hat{\alpha}')^2 = 1 - O_p\left(\frac{1}{n^{\frac{1}{2}+\varepsilon}}\right)$

iii) $\hat{\alpha}'^{\,2} = O_p\left(\frac{1}{n^{1+2\varepsilon}}\right)$

Since

$$\hat{\theta}_2' - \bar{Y}_N = \bar{X}_N\left[\hat{\alpha}'(B-R) + \hat{\alpha}'\frac{\sum x_i u_i}{\sum x_i^{\,2}} + (1-\hat{\alpha}')\frac{\sum s_i}{\sum x_i}\right] ,$$

$$(\hat{\theta}_2' - \bar{Y}_N)^2 = \bar{X}_N^{\,2}\left[\hat{\alpha}'^{\,2}(B-R)^2 + \hat{\alpha}'^{\,2}\left(\frac{\sum x_i u_i}{\sum x_i^{\,2}}\right)^2 + (1-\hat{\alpha}')^2\left(\frac{\sum s_i}{\sum x_i}\right)^2\right.$$

$$+ 2\hat{\alpha}'^{\,2}(B-R)\left(\frac{\sum x_i u_i}{\sum x_i^{\,2}}\right) + 2\hat{\alpha}'(B-R)(1-\hat{\alpha}')\frac{\sum s_i}{\sum x_i}$$

$$\left. + 2\hat{\alpha}'(1-\hat{\alpha}')\frac{\sum s_i}{\sum x_i}\;\frac{\sum x_i u_i}{\sum x_i^{\,2}}\right] .$$

Omitting terms of probability order greater than 1/n,

$$\bar{E}[(\hat{\theta}_2' - \bar{Y}_N)^2] = \bar{E}[(1-\hat{\alpha}')^2 (\frac{\Sigma s_i}{\Sigma x_i})^2] \doteq (1-\frac{n}{N})\sum^N \frac{(y_i - R_N x_i)^2}{n(N-1)} .$$

XVII. APPENDIX 8

We now state without proof two theorems. Proofs are available in Fuller (13).

Theorem I:

Let $\langle X_n \rangle$ be a sequence of a real-valued k-dimensional random variable such that plim $X_n = X$.

Let g be a real valued function mapping X_n into a R dimensional space. Let g be continuous, except on the set D where $P[X \epsilon D] = 0$. Then plim $g(X_n) = g(X)$.

Theorem II:

Let $\langle X_n \rangle$ be a sequence of k-dimensional random variables with element

$$X_n^{(j)} \qquad j = 1, \ldots, k$$

and $g_n(X_n)$ be a sequence of measurable functions. Let $\langle s_n \rangle$ and $\langle r_n^{(j)} \rangle$ be k+1 sequences of positive numbers.

If

$$X_n^{(j)} = O_p(r_n^{(j)}) \qquad j = 1, \ldots, t$$

$$X_n^{(j)} = o_p(r_n^{(j)}) \qquad j = t+1, \ldots, k$$

and if for any nonrandom sequence $\langle a_n \rangle$ there exists an N such that for $n > N$

$$g_n(a_n) = O(s_n)$$

whenever

$$a_n^{(j)} = O(r_n^{(j)}) \qquad j = 1,\ldots,t$$

$$a_n^{(j)} = o(r_n^{(j)}) \qquad = t+1,\ldots,k$$

then for $n \geq N$,

$$g_n(X_n) = O_p(s_n) \quad .$$